Clinical Reasoning Development in Medical Students:

An Educational and Transcultural Comparative Study

Ana L Sergio Da Silva, BSc.Ed(Hons),PGDipHR

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

2013

Abstract

Clinical reasoning research has concluded that experts use less, but more selective, knowledge in a more efficient way, based on the construction of schema, scripts and other representations of the relation between signs, symptoms and diagnoses, derived from their experience. However, this conclusion does not help Medical Schools to decide which pedagogical strategies should be adopted to foster clinical reasoning in undergraduates. This study aims to investigate how medical students, approach clinical cases and the impact of three types of curriculum upon their clinical reasoning.

Two studies were carried out. The first analysed 60 hours of Problem-Based Learning sessions using electronic content analysis and corpus analysis. A second used a cross-sectional approach assessing and comparing students' clinical reasoning in three different medical schools (Derby, Nottingham and Coimbra) based on a Clinical Reasoning Test (CRT) developed and validated for the purposes of this research.

The analysis of the PBL sessions indicated that early contact with clinical cases might favour students' encapsulation of knowledge. First year students use more words, are more descriptive and make significantly more use of explanations. Second year students are more focused using less words, focusing more on the biomedical sciences aspect of the cases and engaging more in questions.

The comparisons between different medical curricula show some differences between groups, at the entry to practice level in favour of the PBL and the integrated curricula. However, at the graduation level only small differences remain between the groups. Clinical exposure has a significant impact in improving students' clinical reasoning, with differences in exposure time between curricula possibly accounting for such results. Additionally, differences in the strategies used to approach the cases were noted students from the traditional curriculum seem to be waiting until all information is displayed to make a decision, while their peers from other curricula seem to be more willing to make decisions based on initial patient's information. No significant correlations with knowledge about the cases, or confidence on the diagnosis were found; possible reasons for these results will be discussed and implications for curriculum development and future research highlighted.

Acknowledgements

"Begin at the beginning, […] and go on till you come to the end: then stop"
(Lewis Carroll, Alice in Wonderland, Chapter XII)

The research here presented was made possible by a doctoral grant awarded by the Portuguese Foundation of Science and Technology (FCT) (SFRH / BD / 39670 / 2007) and funded by POPH - QREN in cooperation with European Social fund and Portuguese Ministry of Science and Technology.

This research was made possible due to the contributions of many people, to all many thanks. I wish to show my deep appreciation and gratitude to my supervisors' their generosity; expertise and commitment made the work here presented possible and my learning experience both a professional and personal invaluable. Discussions with them have constantly helped to shape my own thoughts.

Professor Reg Dennick, my supervisor, by his crucial contribution this work, for his supportive nature and energetic enthusiasm, for all the advice, help and the important lessons and for being an role model of excellency in educations.

Professor Jose Pereira Da Silva, mentor and co-supervisor of this project, without which this work would not be possible, for introducing me to medical education, for his tireless support as a mentor, as friend and a outstanding inspiration both personally and professionally.

Professor Filomena Gaspar and Dr. Danny McLaughlin, co-supervisors, for all the useful advises and support to this project and to me as a learner.

And to all that I had the pleasure to work along side and to learn from at the Medical Education Unit (UoN), at the University of Nottingham and at the Graduate Entry Medical School in Derby (UoN).

I wish to express my gratitude to the outstanding institution that is at Faculty of Medicine of University of Coimbra, Portugal for the opportunity to start my career in medical education and to its director, at date Professor Santos Rosa, by the opportunity to collect data in this institution during the present research. Furthermore, I wish to express sincere acknowledgement and gratitude to Professor Catarina Resende De Oliveira and Professor Julio Leite for supporting the grant application and enthusiasm for this project.

I also want to express my appreciation to the Medical Education Unit University of Nottingham director and staff for welcoming and supporting me during these last four years. To Dr. Tim Heywood, Dr. Swe Khin-Htun and Dr. Nick Bailey for all the input on the clinical cases and

clinical reasoning test developed; to Dr Simon Wilkinson, Mr Anthony Brown and Mrs Anne Rockcliffe for all the IT and touchstone support.

The work here presented was only possible due to all the volunteers, students and facilitators who took part I this research. To all those my most deep gratitude for their time and willingness to share their experiences, this research could not have been possible without them.

I couldn't finish this section without expressing my most sincere gratitude to my family who are and will always be my *roots*, without their support, help and love this journey would not have been possible.

Finally, to Pedro for all support and understanding throughout the 'ups and downs' of this journey. Thank you so much for being part of my live.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

''Science is a special kind of story-telling with no right or wrong answers, just better and better stories''. (Mary Budd Rowe, a Stanford professor of science education, in Norman, 2004)

A. Introduction

Clinical reasoning – that is, the ability to mobilise, interpret and efficiently manage medical and biomedical knowledge and doubt in a clinical context with the purpose of solving clinical problems – is one of the most crucial competencies of the expert physician (Boshuizen *et al.,* 1997; Ericsson, 2007). This complex process has motivated extensive and prolific research during the past four decades (Norman, 2005; Elstein, 2009). Understanding how doctors think*,* how in the face of uncertain situations they identify cues, process information, search their extensive medical knowledge, and retrieve and apply relevant knowledge leading to positive outcomes for patients has always puzzled both the medical and the research community (Schuwirth, 2009).

However, most of the studies in this area have been designed to understand the expert clinicians thought processes, and how they arrive at diagnoses, mainly by comparing them with novices (Elstein, 2009; Rikers & Verkoeijen, 2007) and, in a few studies, with medical students (Eva, 2005a).

Research over the past 40 years has concluded that experts use less, but more selective, knowledge in a more expeditious way, based on the construction of schemas, scripts and other representations of the relation between signs, symptoms and diagnoses, derived from clinical practical experience (Norman, 2005; Elstein *et al.,* 1978; Schmidt *et al.,* 1990; Charlin *et al.,* 2000; Norman and Eva, 2005; Charlin *et al.,* 2007). Research has also shown that there is a strong link between medical knowledge and clinical reasoning (Patel *et al.,* 1990; Woods *et al.,* 2007; Boshuizen *et al.,* 1992) and that experience has an invaluable effect (Schmidt *et al.,* 1990; Norman & Eva, 2005; Norman, 2006).

"Expertise in clinical reasoning is neither mastery of analytical rules nor accumulation of experience, it is both." (Norman, 2000 p. S132).

On the other hand, research also suggests that the ways students learn might have an important influence on the development of clinical reasoning (Anderson *et al.*, 2008; Boshuizen *et al.*, 1992; Medin *et al.*, 1984; Patel, *et al.,* 1993; Schmidt & Boshuizen, 1993a, 1993b; Schmidt *et al.,* 1987). Student-centred, active learning, case-based teaching and assessment, and especially, problem-based learning are all devoted to the development of clinical competency (Cooke *et al.*, 2007). The

difference, in comparison to classical Flexnerian[1] teaching, lies in the ability to use knowledge effectively, and to integrate knowledge more efficiently; that is, to engage in clinical reasoning. That in turn must be seen as one of the most important goals of Medical Schools and regulatory bodies (Cooke *et al*., 2007; GMC, 2009), and for the last 100 years or more has been the driver for highly fruitful innovations in medical education practices and curriculum development (Cooke *et al.*, 2007).

Nevertheless, research into the impact of different curricula on clinical reasoning has been sparse (Eva, 2005a), and not always consensual. A literature search shows a remarkable scarcity of studies dedicated to the identification of factors affecting the development of clinical reasoning in undergraduate Medical Education (Anderson, 2006; Groves, 2002). Additionally, the majority of research aimed at understanding or comparing the effects of curricula on the development of clinical reasoning has looked at problem-based learning curricula versus traditional curricula (Goss *et al.*, 2011; Neufeld *et al.*, 1981; Rich *et al.*, 2005). This has ignored other widely used types of curriculum such as the integrated model. Only one study by Schmidt *et al.*, in 1996 (Schmidt *et al.*, 1996) considered this type of curriculum in addition to the problem-based and traditional curricula. Their study concluded that, although in the final-years PBL and integrated curricula students showed no difference, in the early years of their course (2nd and 3rd) the students from the integrated curriculum demonstrated better clinical reasoning. This led the authors to call for more research into the matter (Schuwirth, 2009; Rikers & Verkoeijen, 2007; Neville, 2009). Furthermore, there is a clear lack of developmental or longitudinal studies aimed at understanding the development of clinical reasoning at the undergraduate level. To this author's knowledge there is only one, a study conducted more than 30 years ago by Neufeld and colleagues (Neufeld *et al*., 1981), following a small sample of 22 students from McMaster Medical School in yearly intervals from undergraduate to postgraduate level. These authors video recorded the participants in a simulated patient encounter followed by a stimulated recall to record their thought processes. Data was then coded and analysed in order to characterise the clinical reasoning

---

[1] Describes a medical curriculum with a strong component of basic sciences, where, usually the three first-years are dedicated to basic sciences and the remaining years are dedicated to clinical practice training.

process and measure the outcome. Students results were then compared with the results form a doctor criterion group. They concluded that these students' clinical reasoning processes seemed to be relatively constant, when compared with the doctors group, from medical school to entry into practice, although an improvement was seen in diagnostic and measurement outcomes due to education. They concluded that more research is needed in order to provide better understanding of the development of this process (Neufeld *et al*., 1981).

In summary, although some research has been conducted into the impact of different curricula on the development of clinical reasoning, much more is needed in order to assist Medical Schools in the difficult task of deciding which strategies should be adopted to foster clinical reasoning in undergraduate students, and to make sure that this mental process continues to develop throughout their professional lives.

The present research project aims to address this lack of information and add to the body of knowledge within this research field (Tashakkori & Teddlie, 1998) by conducting two complementary studies, one aimed at the process of clinical reasoning in a PBL environment and one aimed at the products of clinical reasoning, comparing three different medical curriculum models.

The first study is aimed at understanding the process by which medical students at a very early stage of their curriculum – without extensive medical knowledge – approach and explore clinical problems during PBL sessions. Around 60 hours of PBL sessions from first and second-year groups were filmed, audio-recorded and analysed using a methodology adapted from linguistics: corpus analysis and electronic content analysis. In these sessions, students are presented with clinical cases to discuss in a purposeful and educationally planned way. This allows identification of key learning areas that provide direction to students' self-directed learning throughout that week. Recording this process provides a unique opportunity to gain insight into how students deal with such cases, how they mobilise their acquired knowledge, how they share that knowledge with the group via explanations and questions, and what connections are established between acquired and new knowledge in the context of that clinical problem. Electronic content analysis and corpus analysis are extremely useful methods to analyse such a process in a feasible and objective way. Electronic content analysis allows the identification of major themes in the students' discourses by identifying clusters of words with related meanings and mapping the text according to the frequency of those clusters, automatically based on an extensive

database of dictionaries and thesauruses. Corpus analysis tags every word in the texts according to its semantic and syntactic value, based on many decades of study of the use of the English language in different contexts. This method allows for words to be classified into semantic and syntactic categories and subcategories, which can then be searched, combined, and retrieved, based on the aims of the research. Indicators such as absolute and relative frequencies of predefined categories (for example words related with the body and individual or cause/consequence), sub-categories (e.g. Anatomy), or even single words can be used to perform statistical comparisons between texts. These comparisons can be performed automatically by corpus analysis software using log-likelihood statistics (an adaptation of the Chi-square method), but the data from corpus analysis can also be exported into common statistical programmes such as SPSS 19.0, for further statistical analysis. This was the first time these combined methodologies were applied to the study of transcripts of PBL sessions in clinical reasoning research, as an alternative to more traditional qualitative methods of analysing video and audio recordings of students interacting with clinical cases.

A second study, compared the outcomes of clinical reasoning from two cohorts of students, one with comparatively little clinical exposure, and one with much more (near graduation), in three different medical schools, each representing a common type of curriculum in medical education (a Problem-Based Learning graduate entry course, a Systems Based Integrated curriculum, and a Traditional curriculum, respectively). This study aims to understand the differences in students' abilities to engage with clinical cases, investigating possible differences in the strategies they use, and to see how these relate with the curriculum characteristics. In order to do so, we will make use of the unique opportunity to compare students from a traditional medical school with a predominantly passive teaching tradition, in Coimbra, Portugal; students in a medical school delivering a more contemporary systems-based integrated curriculum, the BMedSci/BMBS undergraduate course in the University of Nottingham, UK; and a graduate entry medical school which has adopted the problem-based learning strategy, also part of the University of Nottingham, UK. In order to do this, a clinical reasoning instrument was specifically developed, the Clinical Reasoning Test, CRT, validated and used for the purposes of this study, in order to capture differences in the overall performance between the cohorts, allow for further identification of differences in the strategies used and the

amount of information needed before making diagnoses. The CRT is a theory-driven, cased-based, flexible instrument that incorporates some of the characteristics of other instruments used to assess clinical reasoning, recombining them in order to maximise their potential to assess clinical reasoning at the undergraduate level.

For such a complex process as clinical reasoning, one needs to make use of a plurality of methods and perspectives, in order to provide a more accurate view of the educational factors that have an impact on its development. Only by doing so can the ultimate goal of the present research be achieved, to provide significant contributions towards the on-going debate on clinical reasoning and the effects of different curricula on the development of this high order mental process in undergraduate medical education. The next section will provide a description of the philosophical approach underpinning the current research purposes, outline the methodological choices, research questions and methods used.

B. Philosophical Approach

B.1 Research Paradigms

The word paradigm as been popularised in the field of science by Kuhn in his book *The structure of scientific knowledge*. In his work Kuhn has defined paradigms as a disciplinary matrix that gather the shared understandings of the world, and nature of knowledge within a scientific community and/or field (Kuhn, 1960, 2000). However some pointed that Kuhn himself has used this word in more them 20 different ways through his early work (Masterman 1970 in Morgan 2007). In epistemology this has led to an extensive debate on the nature of knowledge and knowledge creation, and especially how these shared systems of values, beliefs and practices relate with scientific revolutions and breakthrough moments in modern science (Tashakkori & Teddlie, 2003). With one of the most famous moments being the debate between Kuhn and Popper in the *Criticism and The Growth of Knowledge* conference organised by Imre Lakatos during the 1970's (Lakatos, 1970 in Rowbottom, 2011) (for a more detailed view on this debate see Rowbottom, 2011). These discussions transcend the scope of the present thesis.

Importantly for the present research is the fact that education, as many of the social sciences, is considered a non-paradigmatic research field, as opposed to other sciences such as physics or mathematics, where a single paradigm is adopted at any one time as holding the key to the creation of knowledge and intelligibility of reality (Alexander, 2006).

In education and social sciences several, competing paradigms are accepted to co-exist in the research time-space. The realities these sciences try to understand are so complex and multifactorial that is now accepted can only be truly understood by multiple views (Creswell, 2003).

Medical Education is a research field of education, by the nature of knowledge, object of study, and research practices, inheriting its non-paradigmatic nature. Bunniss and Kelly (2010) for example identify four main types of paradigms common in Medical Education similarly to other areas of research are : positivism, post-positivism, interpretivism and critical theory. Each of these represents a different view on the nature of reality (ontological believe), nature of knowledge (epistemological beliefs) and nature of research (methodological beliefs). All of these co-exist within this field of research, demanding from the researchers to make an

effort to justify the paradigm underpinning their research in order to allow others to understand their research and design choices (Alexander, 2006; Creswell, 2003).

B.2 Pragmatism Approach Or Paradigm

Pragmatism as a paradigm is not new in education, works such as of Dewey or James are examples of its application to education. Pragmatism as an approach or a paradigm underpinning research, can be defined as focused on "consequences of actions", "problem-centred", "pluralistic" and "real-world practice oriented" (Creswell, 2003 p.8). These characteristics convene a simple idea that methodology and all research decisions should be aligned with the research purposes, and the questions posed. This fact increases the need for justification of researcher's choices and demands increased responsibility to provide explanations for the rational adopted in the research. The inherent flexibility of this paradigm is not without criticisms form supporters of other paradigms. If used incorrectly often can be seen as an all-encompassing justification for a certain 'philosophical anarchy', whereby the researchers fail to express the philosophical basis of their works (Morgan, 2007). Therefore, it is of concern to briefly outline the assumptions that underpin pragmatism and how those have influence the present research and how it differs from others commonly use paradigms in medical education.

Contrariwise to the positivist and post-positivist paradigms, pragmatism does not conceive the nature of reality to be a single static and unchangeable overarching truth ready to be uncovered by research. On the other hand, pragmatism does not assume, like interpretivism that that all reality is subjectively constructed in the interaction between the researcher and the object of research. Pragmatism accepts that there is an 'external reality' that is translated into knowledge through the lenses of the researchers. In this sense, knowledge driven from research is only a better or more adequate explanation for outcomes rather 'universal truth'. Here, "truth being what works at the time" (Maudsley, 2011 p.e95).

As Creswell (2003) described our aim is "real-world practice oriented", as we aim to be able to develop knowledge that can contribute to assist medical schools to promote the development of clinical reasoning of undergraduate medical students. We propose to do so by providing the best possible explanation for the research initially questions posed, based on the outcomes of the present research. We consider those explanations not as absolute laws that can be generalizable across

different contexts and situations, but as new pieces in the puzzle of what is currently known about the development of clinical reasoning in this particular context.

Other of the fundamental characteristics of pragmatism is epistemological relativism (Maudsley, 2011). On this aspect pragmatism is 'between' those paradigms that define knowledge as being an "value-free" description of an objective reality (e.g. positivism), and those that considered knowledge to be always a subjective co-construction (e.g. interpretivism, critical theory). Pragmatism accepts both objectivity and subjectivity are important for knowledge creation. Additionally for pragmatists knowledge depends on researchers' interpretation of an 'external reality' therefore it cannot be entirely value-free (Creswell, 2003).

This epistemological assumptions impact directly on the approach to research methodology and methods used. Contrary, for example, to positivism that seeks to deduct laws and find causality, pragmatism accepts the search for causality is a potential research purpose although is an elusive one. Here the research purposes are defined by the researcher based upon the problems identified ("problem-centred"). The use of both inductive and/or deductive is accepted depending on the purpose chosen (see Tashakkori, & Teddlie 2003 for a taxonomy of research purposes).

Hence pragmatism assumes a "pluralistic" position with regard to methods used (Creswell, 2003). Differing from other research paradigms, where the nature of its ontological and epistemological assumptions demand that specific methods are adopted, in pragmatism it is the responsibility of the research to ensure this coherence. It assumes that both qualitative, quantitative are equally valid ways to research as long as consistent with the research questions and problems to be investigated ("consequence of action"). In fact, mixed-methods research (MMR) is often situated within this paradigm (Creswell, 2003). This view contrast with for example positivist or post-positivist paradigms that assume that can only be researched using objective quantitative methodologies that will allow research findings to be generalisable across context and populations.

In the current research we assume that development of clinical reasoning as an 'external reality', that is something that is not depend of the current research and that we aim to investigate focusing on specific identified questions ("problem-centred"). These questions guided our choice to design a mixed-methods research, combining two studies adopting different methods ("pluralistic"). One aimed at

understanding how early stage medical students discuss clinical cases in a PBL setting and a second aimed at understanding the impact of different curricula in the development of medical students clinical reasoning. Finally outcomes of both studies are discussed and juxtaposed in order to develop answers for the questions initially posed.

However we assume the outcomes and answers emerging from these studies are not independent of our methodological choices and the contexts where the research was developed. Also we assume that answers presented are to be understood as a possible interpretations of the outcomes based upon the researchers' own lenses ("consequence of actions").

In summary, the pragmatism approach emerges as a practical alternative to the metaphysic "paradigm incompatibility" discussion motivated by the need to choose the best possible way to answer the research questions (Maudsley, 2011). This believe resonates with our current views and believes about the nature of knowledge and the role of research in education and in medical education.

C. Methodology

C.1Mixed-Methods Research (MMR)

The present research adopts a mixed-methods methodology (MMR). This type of methodology has a long tradition within the social sciences and education with some works such as Campbell and Fiske in 1959, Sieber 1973, Denzim 1978, Bryman 1988 and more recently Tashkkoiri and Teddlie 2003 or Creswell 2009 to mention just a few of the most commonly cited examples (Creswell & Clark, 2007; Heyvaert et al., 2011; Creswell, 2009; Teddlie & Yu, 2007; Tashakkori & Teddlie, 1998; Niglas, 2004).

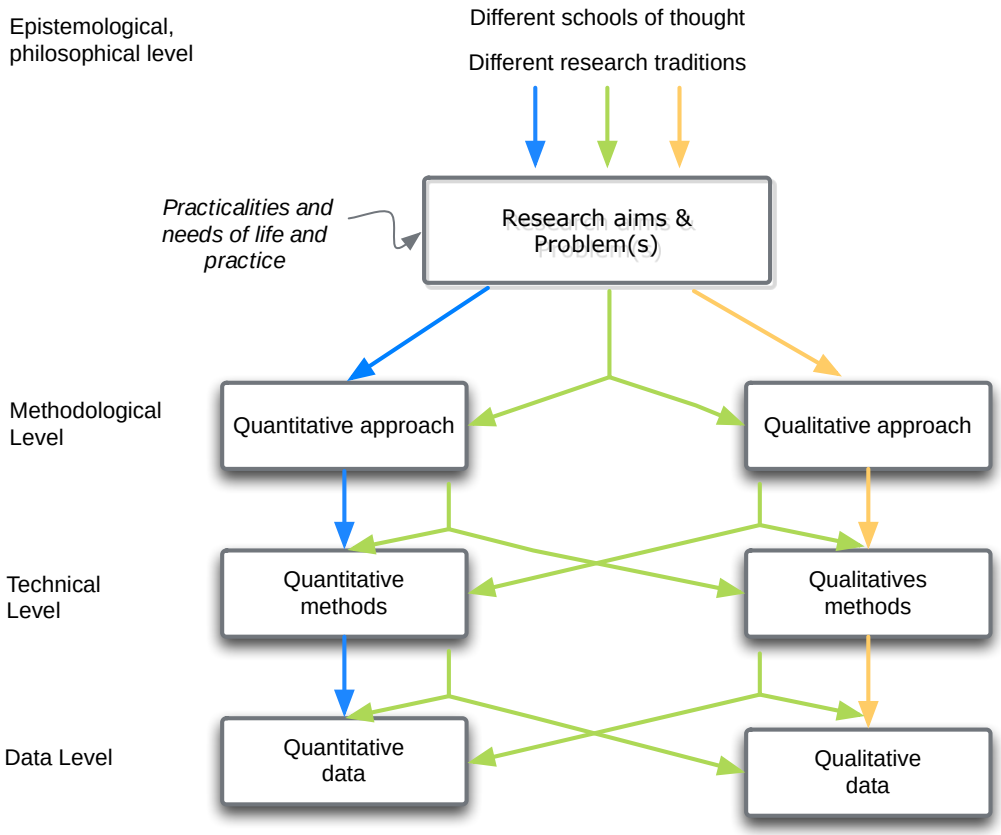
In recent years the use of MMR has become increasingly more popular in medical education, with some advocating that this methodology has potential to be popular in this research field but it requires careful use and adequate understanding the criteria for use and for combination of the qualitative and quantitative elements (Maudsley, 2011; Zhang, 2011).

Mixed-methods research (MMR) is usually based upon the pragmatism paradigm or approach and assumes the research problem as the centre of the research design. In this type of research qualitative and quantitative methods, data and analysis techniques can be combined in order to provide the best possible understanding of the problem. Within the mixed-methods research both elements (qualitative and quantitative) can be combined in many different ways. Because in this research methodology more freedom is given to the researcher to decide what best suits the research problem there is a diversity of possible research designs (Tashakkori & Teddlie, 2003). Heyvaert et al. (2011) suggests a classification of 18 types of MMR based upon the weight that each of the elements assume in the research, the level of integration between both elements and the temporal orientation or timing of that integration (e.g. concurrent or sequential), all being equally considered in MMR research. This is clear in Figure 1.1 where qualitative, quantitative and mixed-methods perspectives are presented, with the last combining elements of the two previous ones at different levels (Niglas, 2004)

.

Figure 1.1: Qualitative and qualitative methodologies in mixed-methods research according to Niglas (Niglas, 2004) classification.

Research levels



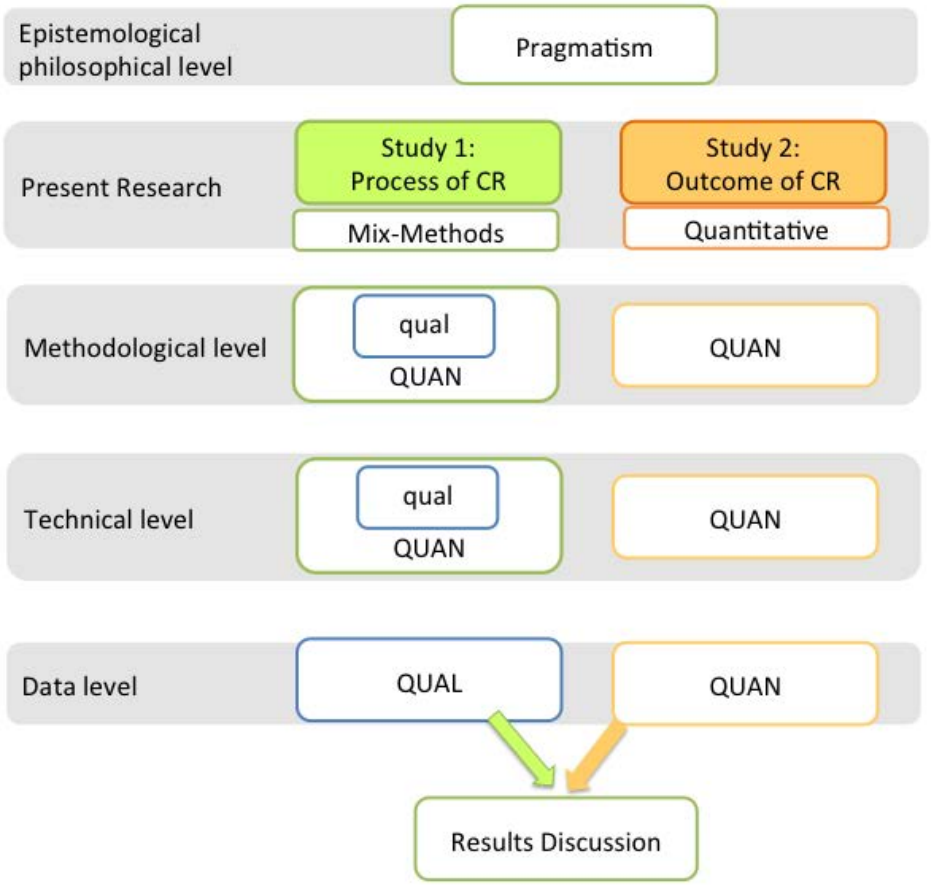
Mixed-methods in the present research

Educational research is by nature multifactorial and shaped by the interactions between several different factors (Conrad, 2006); a holistic approach and multiple methods are required in order to understand it (Berliner, 2002). Moreover, clinical reasoning is also a highly complex construct. Clinical reasoning cannot be directly measured; therefore we rely on indirect measures and inferences made from those. Additionally, no single definition or single instrument has been agreed upon in the medical education community as the 'gold standard' for assessing clinical reasoning

(Norman, 2005). Neither has a clear nomological network[2] in which to frame clinical reasoning been produced (Salkind, 2010; Borsboom, 2008; Borsboom, 2005).

In the present research (see Figure 1.2), corpus analysis and electronic content were used to analyse aspects of transcribed video recordings of PBL sessions analysis using both qualitative and quantitative data, with a stronger emphasis on the quantitative reporting of the results. The results of that study were discussed along with the results of the comparative quantitative study in order to achieve a better understanding of possible differences in way impact of different curricula on the way medical students develop clinical reasoning.

Figure 1.2: Application of Niglas (2004) classification to the present research



2. A nomological net or network is "a pattern of relationships among variables in which the constructs are embedded". Regarding clinical reasoning, although some variables such as medical knowledge, can be identified, there is not clear identification of the relationship, we only know they are linked, not how or how strongly. The same case happens with reasoning; so far no study has proven beyond doubt the relationship between generic types of reasoning and clinical reasoning. (Salkind 2010)

Research Aims and Questions

Study aims

Understand the process of development of clinical reasoning at early stages of the medical undergraduate curriculum.

Understand the impact of different curriculum types on the development of students' clinical reasoning strategies at the undergraduate level

Understand the impact of clinical experience/exposure on students' clinical reasoning strategies.

Research questions

Table 1.1 provides an overview of the research questions identifying for each the research aims, instruments used and the data collection procedures, providing an integrated summarised view of all the research that will be described and discussed in the following chapters. These questions were arose both from the researchers interest in the subject, after some years of experience working with medical students, and from the review of literature on the subject presented in the following chapter.

D. Rigor And Quality Considerations

The criteria for rigour and quality adopted for each of the studies will be presented in each of the methodologies sections (study one-Chapter 4 and study two-chapter 5). Additionally we hope to demonstrate in the present thesis how the work now presented was underpinned by the criteria for quality in educational research suggested by Eisenhart and Howe (1992): *"cogently developed; competently produced; coherent with respect to previous work; important and ethical; and comprehensive"* (Eisenhart &Howe, 1992 in Nigla 2004 p.23).

E. Structure And Purpose Of This Thesis

This thesis seeks to contribute to the understanding of the development of clinical reasoning at undergraduate level. It achieves this by looking at how students deal with clinical cases early in their curriculum, and how different educational factors – mainly the curriculum type, and exposure to clinical practice – impact upon the students' clinical reasoning process.

A necessary background context is provided in Chapter Two, where a review of existing literature on clinical reasoning will be presented.

Chapter three presents the educational context in which the study takes place, describing and comparing the curricula in the three medical schools.

Chapter four will describe the adaptation of corpus analysis to the study of medical and PBL contexts and the methodology used in study one: analysis of PBL sessions. Chapter five. Will present the results, discussion and limitations of this study.

Chapter six describes the development and validation process of the Clinical Reasoning Test (CRT) and the methodology used to compare clinical reasoning strategies and outcomes of two groups of students from three different medical schools, each representing a common type of curriculum in Medical Education (study two).

Chapter seven presents the results, discussion and limitations of this study and the application of the CRT to the initially defined research questions.

Chapter eight presents the overall discussion of the findings of the present research, implication for curricula development, further research and conclusion.

Table 1. 1: Research design overview

Aim	Research method	Research questions (summary)	Instrument/Tool	Data collected
1. Understand the process of development of clinical reasoning at early stages of the medical undergraduate curriculum.	Corpus analysis and complementary electronic text analysis (study one). Chapters 4 & 5	Can the development of early basis of clinical reasoning be further understood by analysing the discussions that take place during PBL sessions? What is the relationship between the session's discussion and the cases diagnosis as presented in the facilitators' guides? How do the students explore the cases? (Strategies) Are there any differences in questions, reasoning and explanations between sessions, cases, modules, and years?	Electronic text analysis software (Leximancer); Corpus analysis software: Wmatrix2 Statistical software: SPSS.19	Video and audio recording of PBL sessions
2. Understand the impact of different curriculum types on the development of students' clinical reasoning strategies at the undergraduate level. 2.1. Definition of the context of the study	Context of the study: Analysis of the curricula. Chapter 3	What is the role of clinical reasoning on regulatory bodies documents in two countries? What are the main differences and similarities between the three types of curriculum: In their stated aims/teaching strategies/assessment structure? Opportunities for clinical reasoning	Curriculum evaluation framework (adapted from Harden *et al.*, 1984; Harden, 2000).	Publicly available primary sources.

		development? Clinical exposure and clinical practice opportunities?		
2. Understand the impact of different curriculum types on the development of students' clinical reasoning strategies at the undergraduate level. 2.2. Comparison of students outcomes/results. 3. Understand the impact of clinical experience/exposure on students' clinical reasoning strategies.	Cross-sectional comparative study (study two). Chapters 6 & 7	How can clinical reasoning be assessed at an undergraduate level? What is the impact of the three types of medical undergraduate curriculum on student's Clinical Reasoning strategies at the beginning/end of their clinical phases? Are there any differences in student clinical reasoning ability as measured by the CRT? What is the impact of clinical practice/exposure in these differences? How does the students knowledge of the pathophysiology of a disease correlates with students' ability to do a diagnosis of that disease?	Data collection tool: Clinical Reasoning Test (CRT). Analysis tools: Statistical software SPSS 19.0	Answers to online test (Touchstone and Survey Monkey).

Chapter 2: Background

"If I have seen further it is by standing on the shoulders of giants."

Isaac Newton (no date)

Chapter summary

The aim of the present research is to understand the educational factors, at the curriculum level that impact on the development of clinical reasoning at the undergraduate level. To do so it is vital to understand the literature on clinical reasoning and curriculum development, in order to inform both questions and directions followed by the present research. This will be presented in the following chapter.

A. Clinical Reasoning Research: An overview

A.1. Clinical Reasoning: A Brief History

"There is a rich ongoing debate about our understanding of the complex process of clinical diagnostic reasoning" (Bowen ,2006)

Clinical reasoning, that is the way in which clinicians solve patients' problems searching into the extensive databases of knowledge in their minds, in a fast and accurate way has always puzzled the medical community and interest in it can be traced to the beginning of modern medicine. Sir William Osler (Osler, 1899) for example, in some of his writings mentions the importance of developing one's reasoning as a way to become a competent physician (Osler, 1909). Clinical reasoning was seen as something fundamental but at the same time somehow unexplainable and mystical, a Holy grail of medicine for centuries (Schuwirth, 2009).

Throughout time a vast body of research has accumulated around this theme, although little consensus has been achieved (Norman, 2005). Two main reasons are highlighted: One is the dispersion of research across many different areas (from sociology to artificial intelligence) with poor collaboration and communication between them (Norman, 2005). Second is the lack of consensus on the fundamental characteristics of clinical reasoning (Bowen, 2006; Patel *et al.,* 2008; Norman, 2005).

For the purpose of this thesis we will focus on the medical education literature, making, when necessary, connections between research here and the research published in other areas.

Research programmes of Elstein and Bordage in the 1970s and Schmidt and Vilma Patel's in the 1980s and early 1990s and the McMaster University since late 1980s to the present date, have shaped clinical reasoning research and were responsible for most of what is now known about clinical reasoning assessment, development and its relationship with other areas of medical education (Ericsson, 2007; Elstein *et al.,* 1978; Coughlin & Patel, 1987; Coughlin & Patel, 1985; Eva & Norman, 2005; Norman *et al.,* 1989; Norman, 2005).

In the late 70s Elstein *et al*. (Elstein *et al.,* 1978) published what became a classic book on the study of medical expertise and clinical reasoning *Medical Problem Solving: an Analysis of Clinical Reasoning*. These authors set out to replicate the work

done at the time in the field of chess, mainly by de Groot (DeGroot, 1946), in the field of medicine. These studies, as others at the time in cognitive psychology (Wiley & Jee, 2011), were based on the assumption that superior performance would be connected with a more developed generic problem-solving skill (Norman, 2005). This concept of heuristic problem-solving ability was conceptually close to the general intelligence factor (g factor theory) proposed by Spearman in the beginning of the 20th century. The aim was to characterise superior (expert) performance in medical problem solving, the diagnosis of, and treatment of patients (Elstein *et al.,* 1978; Ericsson, 2007) by identifying representative tasks. Such tasks would be simultaneously accurate representations of clinical practice context and challenges, in other words opportunities for experts to demonstrate their superior performance (Ericsson, 2007). It is not difficult to understand how this task is much simpler in chess, then in medicine (Shanteau, 1992).

A chessboard will always have the same configuration, chess does not deal with uncertainty, laws do not change making this a much more certain game then medicine, where decisions have to be made in an uncertain context (Elstein & Holmes, 1981; Ericsson, 2007; Elstein, 2009), with many more variables in every "move".

The selection of expert clinicians with a high performance in diagnosis and treatment of patients was also a cornerstone of Elstein's work (Elstein *et al.,* 1978). The author compared the performance of expert clinicians with "normal" clinicians on representative tasks (Eva *et al.,* 2002). These tasks would simulate a patient consultation with the aid of actors as patients, which were video recorded, clinicians would be asked to watch their own performance and explain what they were thinking during the task (Eva *et al.,* 2002). From the comparison between these two groups, no clear differences emerged on diagnostic ability or cognitive processes used (Lyle, 2003).

The introspective methodology used, designated as stimulated recall (Barrows & Feltovich, 1987; Lyle 2003), was criticised by several authors (Lyle, 2003; Ericsson & Simon, 1998), as leading to the report of false memories. Participants, while watching their performance, would report not so much what they were thinking while doing the recorded actions, but what they thought, at the moment of the recall, they should be reporting, developing a coherent narrative of their own actions (van der Vleuten & Newble, 1995; Elstein, 2009; Groen & Patel, 1985). Additionally also the

statistical analysis and interpretations done by Elstein's in his work were criticised (Norman, 2005).

The inability to find clear differences between the groups, the methodological critiques and the growing trend in cognitive psychology (Ericsson, 2007) opposing the idea of a holist problem-solving skill, led to a major shift in clinical reasoning research. As Ericsson, points out:

"failure to capture individual differences in expert diagnostic performance with either costly high or low-fidelity simulations made the issue of capturing expertise mute and permitted researchers to explore phenomena related to the acquisition of diagnostic reasoning more freely." (Ericsson, 2007 p. 1125).

This led to an increasing interest in knowledge structures as a possible explanation for differences between experts and novices. It was the beginning of "The Age of Memory" soon to be followed by "The Age of Mental Representations" (Norman, 2005 p. 420) in the nineties.

Schmidt and Boshuizen (Boshuizen *et al.,* 1992) shared the focus of Bordage (Bordage, 2007) Patel and Groen (Patel *et al.,* 1986) and Norman and colleagues (Woods et al. 2007a) by recognising that the structure of medical knowledge, its internal organisation and relationship with memory are crucial to explaining clinical reasoning development in medicine rather than a generic "problem-solving trait" (Schmidt *et al.,* 1990). These groups, although sharing some assumptions, adopted a different focus in their research. Schmidt, Boshuizen and Bordage focused on understanding how the knowledge is organised in the minds of experts and how that impacts the clinical reasoning process (Norman, 2005). These were later classified as structure theories of medical knowledge (Schmidt *et al.,* 1990). Norman and Eva, on the other hand, were concerned with understanding the process of solving clinical problems: *"where do the hypothesis come from?"* (as classified by Schmidt & Rikers, 2007), that is processing theories of clinical reasoning (Norman *et al.,* 2007), Patel e Groen group researched both the structure and process (Schmidt & Rikers, 2007).

More recently a growing interest in the effect of reflection and experiential learning on this mental process can be observed. Contemporary papers from Epstein (Epstein, 1999), Mamede and Schmidt (Mamede & Schmidt, 2004) among others (Sys, 2003; Kuipers & Pesut, 2004; Thomas & Goldberg, 2007) may indicate that we are now at the beginning of a new "age" in clinical reasoning research, that we suggest, might be the "Age of reflection".

However, there are still many questions to be answered in this field (Norman, 2005) a greater consensus on a definition and boundaries of clinical reasoning is still a work in progress (Mamede, Schmidt & Penaforte, 2008a; Regehr, 2004). How clinical reasoning develops from undergraduate medical education towards an expert clinician level is still an area where research is needed (Regehr, 2004). Additionally implications from clinical reasoning research in teaching (Bowen, 2006; Croskerry, 2009), the impact of the current curricula models in this mental process (Croskerry, 2009a; Norman *et al.,* 2009; Woods *et al.,* 2007) and how it can be assessed both in experts (Mylopoulos & Regehr, 2007) and in medical students (van der Vleuten & Newble, 1995; Schuwirth, 2009) still represent some of the most prominent questions of the field. These last are especially important to assist Medical Schools in the difficult task of deciding which strategies should be adopted to foster clinical reasoning in undergraduate students and to make sure that this will continue to develop through their professional lives.

A.2. Clinical Reasoning: A Possible Definition

Holyoak and Morrison (2005) observe in the introduction to *"Thinking and Reasoning" a Cambridge Handbook*, that it is not an easy task to provide a clear scientific definition of mental terms that are critically part of everyday language, such as thinking, reasoning or judgment (Holyoak & Morrison, 2005). This is equally true for clinical reasoning (Norman, 2005). Clinical reasoning is part of healthcare specific language; it is embedded into the healthcare professionals' jargon and routinely by many without a reference to a particular common.

Clinical reasoning is often mentioned in the medical education literature, however it is also possible to find concepts, like critical thinking, diagnostic thinking, clinical problem solving, clinical decision making or critical judgement being used as synonyms for clinical reasoning, or as part of circular definitions that can be of little use to research (Simmons, 2010). Although it may seem obvious to those in the medical and medical education field what clinical reasoning stands for, to pursue research in this field it is necessary to have a clear and objective definition of the object of the research (Simmons, 2010). A definition from Kassirer (2010) was used as a starting point in the search for an operational definition of clinical reasoning to use in the present research. According to this author, clinical reasoning can be defined as:

"Clinical cognition [or clinical reasoning] the range of strategies that clinicians use to generate, test, and verify diagnoses, to assess the benefits and risks of tests and treatments, and to judge the prognostic significance of the outcomes of these cognitive achievements." (Kassirer, 2010 p. 1118).

This definition encompasses both the notion that clinical reasoning is a high order cognitive and metacognitive process (Barrows & Tamblyn, 1980; Marcum, 2012) that operates through **all the phases of the clinical cycle** (history, diagnosing, testing, treatment and prognosis). It also highlights the importance of metacognitive processes (Higgs & Jones, 2000; Marcum, 2012; Croskerry, 2009), to judge and reflect on possible outcomes of those cognitive activities as part of the regulation of their own reasoning process (figure.

Throughout over more than 30 years of research into clinical reasoning some distinctive characteristics of this process have been identified, these add to the above definition a more fine-grained image of clinical reasoning (Elstein, 2009; Elstein & Bordage, 1991).

Clinical reasoning is not medical expertise although it is an important part of it. Several authors have pointed out the difference between medical expertise and clinical reasoning (Kassirer, 2010; Schmidt & Boshuizen, 1993; Custers *et al*., 1996a; Elstein *et al*., 1978). According to this view, clinical reasoning is fundamental for expertise in medicine, but being an expert in medicine covers a wide and complex range of knowledge and skills that go beyond the mental process of solving clinical cases (Kassirer, 2010). This clarification is particularly important, in the context of the present research, as our aim is to study clinical reasoning, and not other components of medical expertise, such as for example communication or clinical skills.

Clinical reasoning is not a trait. More than 30 years ago clinical reasoning research tried to identify a generic ability to solve clinical problems, or engage in an effective clinical reasoning process (Kassirer, 2010). This was a view shared by many in the cognitive psychology field at the time (Wiley & Jee, 2011; Carver & Klahr, 2001). However, further research, both in psychology's wider field and in clinical reasoning research has demonstrated that in fact it is not possible to identify such generic problem-solving abilities, as the ability of individuals to solve problems is highly dependent on the context, characteristics of the problem and the individual knowledge of both (Boshuizen, 2003; Dory *et al.,* 2009; Splinter *et al.,* 2007; Brooks *et al.,* 1991; Norman *et al.,* 2006).

In clinical reasoning research Elstein's attempts at defining a generic clinical problem solving ability in experts led to an interesting conclusion that success on one problem is not a strong predictor of success in a second problem (Elstein *et al*., 1981)[3]. The search for a *clinical reasoning trait* that would allow to identify and predict success in medical problems-solving was then abandoned in favour of a view where case, content and context were understood to impact individuals' clinical reasoning ability (Dory *et al.,* 2009).

Case-specificity and context-specificity started to be considered as features of clinical reasoning with very important implications for learning, teaching and assessment (van der Vleuten & Newble, 1995).

In essence all refer to factors that impact on and can help explain the variability on individuals' performances across cases, with different designations reflecting different possible causes for such variability (Norman *et al*., 2006).

Case-specificity is used simply to refer to the fact that individual performance varies between cases without adding any attempt to identify an explanation for that variability (Dory *et al.,* 2009).

Context-specificity means that the differences in a clinician's performance between cases could be explained by the context in which those cases are learned and/or presented. The effect of context on learning and retrieval is a long-known effect of human behaviour and has been demonstrated by many in memory and psychology studies (Eva, 2003). A classic study by Godden & Baddeley in 1975, demonstrated that recall of words is improved if the context of learning was the same as the context of recall as the physical environment provides cues that impact on the recall processes (Godden & Baddeley, 1975). Therefore, when the context of learning and context of application are similar (e.g. clinical environment) then transference and application of previous examples would be facilitated (Colliver *et al.,* 1990; Eva *et al.,* 1998).

Additionally the **clinical context, situational and ecological factors** in which the encounter is embedded impact the reasoning process, that is clinical reasoning is a *"multifactorial and context-dependent"* (Ajjawi & Higgs, 2008). For example, patient encounter in an ambulatory setting is clearly different from an in-patient ward. These elements are at the centre of the model proposed by Higgs for clinical reasoning of healthcare professionals where the patient encounter and the context of that

[3] For more detail on Elstein's early studies see Section A.A.1 p. 4 of the present chapter.

encounter are at the very core of the clinical reasoning process (Higgs & Jones, 2000). These authors define clinical reasoning as an upwards spiral process in which greater and deeper understanding of the clinical problem is achieved through an interactive process between the clinicians' professional judgment, knowledge and cognitive and metacognitive processes, and the patients' and other healthcare professionals involved in the process (Higgs & Jones, 2000).

Two recent and very different studies one by Durning and one by Sibbald, Panisko & Cavalcanti have also supported this view showing the evidence of the impact of the contextual factors in clinical reasoning and diagnostic performance in doctors. Durning *et al.*, results highlighted the importance of situational and ecological factors, which encompass characteristics of the patient, the encounter and the clinicians own personal factors impacting clinical reasoning performance (Durning *et al*., 2011).

Sibbald and colleagues on the other hand found that not only did the context of the case (given by the researchers) had a significant impact on diagnostic accuracy (increased by 23%,) compared with just the physical examination results (and no context) but was also linked with hypothesis generation and physical examinations' findings interpretation (Sibbald *et al*., 2011).

Literature also suggests that clinical reasoning is **content-specific and knowledge dependent**. Content-specificity was the explanation initially put forward by Elstein's work (Elstein *et al*., 1990). According to this view, differences in performance would be explained by the mastery of knowledge in the area of the case. Clinical reasoning would then be characterised by being highly dependent on an underpinning knowledge base (Dory *et al*., 2009; Norcini, 2003; Mattick *et al*., 2008). This view has been extensively supported by research (Patel *et al*., 2005; Schmidt & Boshuizen, 1993a; Boshuizen & Schmidt, 1992)[4].

Metaphorically one can say mental processes are much like *enzymes* that need a *substrate* (knowledge representations) to act upon. The ability to solve clinical problems effectively is highly dependent on an individual's knowledge about the context of that problem (Castel *et al.,* 2007; Boshuizen *et al.,* 1992). It is the specific nature of medical knowledge that makes clinical reasoning remarkably different from any other type of reasoning (Patel *et al.,* 2005). Structure theories (Schmidt & Rikers,

---

4 For more detail on how knowledge organisation and metal representation studies contributed to a shift on the field of clinical reasoning research see Section A.A.1 p. 4 of the present chapter.

2007) of expertise have suggested explanations for how experts are able to acquire, accommodate and efficiently apply the reasoning process to such large amounts of knowledge (Boshuizen, 2003). Although a consensus is yet to be achieved on the details of the interaction between memory knowledge structures, mental representation and the clinical reasoning process (Patel *et al.,* 1994)[5], it is clear that knowledge organisation structures are critical to clinical reasoning (Patel & Arocha, 1995). Even though it is generally accepted that clinical reasoning is to be considered not a generic trait, but rather a cognitive ability dependent on knowledge, many questions are still to be answered.

Three studies, one from Norman *et al.* (Norman *et al.,* 2006), one by Dory *et al.* (Dory *et al*., 2009) and one by Wimmers and Fung (Wimmers & Fung, 2008) add new elements to previous explanations of case-specificity. Norman *et al.* (Norman *et al.,* 2006) conducted a series of G and D generalizability studies using data from the 6342 graduating medical students on 3 key features sections of the Medical Council of Canada (Norman *et al.,* 2006 p. 618). Each contained a total of approximately 35 cases comprising a written description of the case, followed by a sequence of between 1 to 4 short menu or short answer questions on crucial aspects of the case (key features), e.g. data interpretation, diagnosis and/or treatment.

The authors found that variability in performance across cases was in fact better explained by variability between the items within the cases and that *"Our traditional assumption of case specificity is simplistic. The main source of variability is not due to cases but to the specific items nested in the cases."* (Norman *et al.,* 2006 p. 619). However the authors do not present alternative views or explanations for the item variability.

Dory *et al.* conducted a similar type of a generalisability analysis (G and D) on the results of two cohorts of general practice trainees (total of 227) in 159 extended-matching items sat as part of a compulsory learning day. The items covered four domains (chest, urogenital system, locomotor system, and dermatology) with clinical cases being nested within the topics (Dory *et al*., 2009).

As in Norman *et al.*, this study also concludes that items in the cases rather than the topic of the case topics explain variability in performance and that *"Topical*

---

5 Differences in the way knowledge structures, mental representations and clinical reasoning interact underpin of the differences between clinical reasoning models presented in section: A.A.3 of this chapter.

knowledge does not seem to explain case-specificity observed in our data." (Dory *et al*., 2009 p. 55).

Both studies using different test formats show remarkably similar results, both opposing the initial case-specificity explanation.

Dory *et al.,* raise the hypothesis that structure of relevant knowledge and reasoning strategy used may offer a possible explanation. Trainees could be using pattern recognition with cases they frequently encounter and hypothetico-deductive strategy for not so well known cases, and these strategies were more successful on a case-by-case basis. (Dory *et al*., 2009) This however should not apply so easily to the Norman *et al*., results, as their sample were graduates from medical school, with much less experience, and therefore less likely to be making an extensive use of pattern recognition.

Finally another study from by Wimmers and Fung seems to point to other possible explanation: clinical performance is not a matter of content mastery alone, neither generic abilities (traits), but rather is it's the product of the interaction of both (Wimmers & Fung 2008). These authors analysed the scores of 350 year 3 students' on a clinical performance examination (CPX), mainly looking the sub-score of history taking, physical examination and information sharing and communication involved in five different cases. Using structural equation modelling (SEM), the authors tested the fit of five models, each attributing a different weight to knowledge specific skills and generalisable abilities and its interaction in explaining the students results (Wimmers & Fung 2008). Their conclusions were that:

"Neither a pure knowledge-based perspective nor a pure skills-based perspective on clinical performance seems to predict clinical performance. Content specificity and generalisable skills function together during clinical performance." (Wimmers & Fung 2008, p. 581).

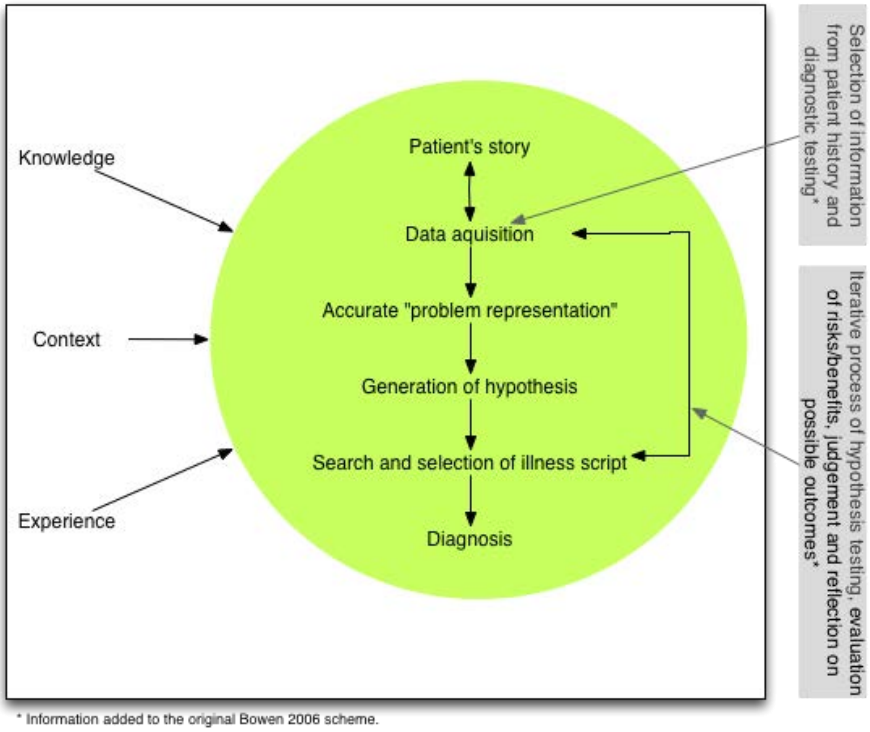
It is important to note that the aim of the study is clinical performance, not clinical/diagnostic reasoning *per se*, although problem solving assumes a major focus on the study discussion, with the authors defending that clinical problem-solving should be better understood as a product of the interaction between domain knowledge and domain-general problem-solving abilities or strategies (Wimmers & Fung, 2008). This explanation may offer an alternative explanation for the results in the two studies mentioned previously with variation in items being a result of variation of generic problem-solving abilities. However, until further research makes

it possible to draw any definitive conclusions, it can only be assumed that knowledge is a fundamental element on clinical reasoning process but alone does not provide a full explanation for differences in performance.

With hindsight of the above mentioned characteristics of clinical reasoning in the present research we define clinical reasoning as the cognitive process, resulting from a combination of cognitive strategies, that allows clinicians to deal with clinical problems, by iteratively generating, testing and verifying diagnoses, establishing treatment plans and prognoses, based upon a previously acquired network of medical knowledge, experience and its interaction with the particular characteristics of the patient, case, context and the situation in which the clinical problem is encountered.

Although recognising that clinical reasoning is not limited to diagnosis and that ecological and situational factors have an impact on this process, due to the scope of the present work, we will focus mainly on diagnostic reasoning elements as shown by Figure 2.1 (adapted from Bowen, 2006 p. 2219).

Figure 2. 1: Key Elements of clinical reasoning adapted from Bowen, 2006. (Bowen, 2006 p. 2219)



A.3. Clinical Reasoning: Models

Clinical reasoning research has not always been consensual when it comes to define how experts solve clinical problems (Norman, 2005; Patel *et al.,* 2005). Although recent advances have been made in order to arrive at more comprehensive explanations of how experts solve clinical cases (Croskerry, 2009a) this is still a work in progress. To understand the current discussions surrounding clinical reasoning, it is necessary to understand models or types of clinical reasoning that have been suggested by the different research groups and authors (Norman *et al.,* 2009). However, key for the present research is that, although important, all of these models are concerned with processes used by experts, offering very little insight on how these processes develop from an early stage to the expert level.

Recently some consensus views have been suggested mainly by Norman (Norman *et al.,* 2009) and Croskerry (Croskerry, 2009a) adapting an idea previously presented by Patel and Laxmisan *et al*. (2005) these will be presented at the end of this section.

According to Schmidt and Rikers, expertise theories can be classified into two groups: Structure theories and Process theories. Structure theories would be concerned with the structures of knowledge[6] that allow experts to acquire, store and mobilise effectively large amounts of information in order to use them in the clinical reasoning process. Process theories on the other hand focus on identifying and providing explanations for the mental processes by which experts solve clinical cases. In other words, they are focused on clinical reasoning (Schmidt & Rikers, 2007). From these, different explanations have emerged through time, such as: hypothetico-deductive reasoning, Backwards vs. Forwards clinical reasoning, Illness Scripts, Non-analytic reasoning, Probabilistic Bayesian reasoning (Elstein & Schwartz, 2002) and, more recently the "universal model" of diagnostic reasoning (Croskerry, 2009a).

Hypothetico-Deductive Reasoning (System 2)

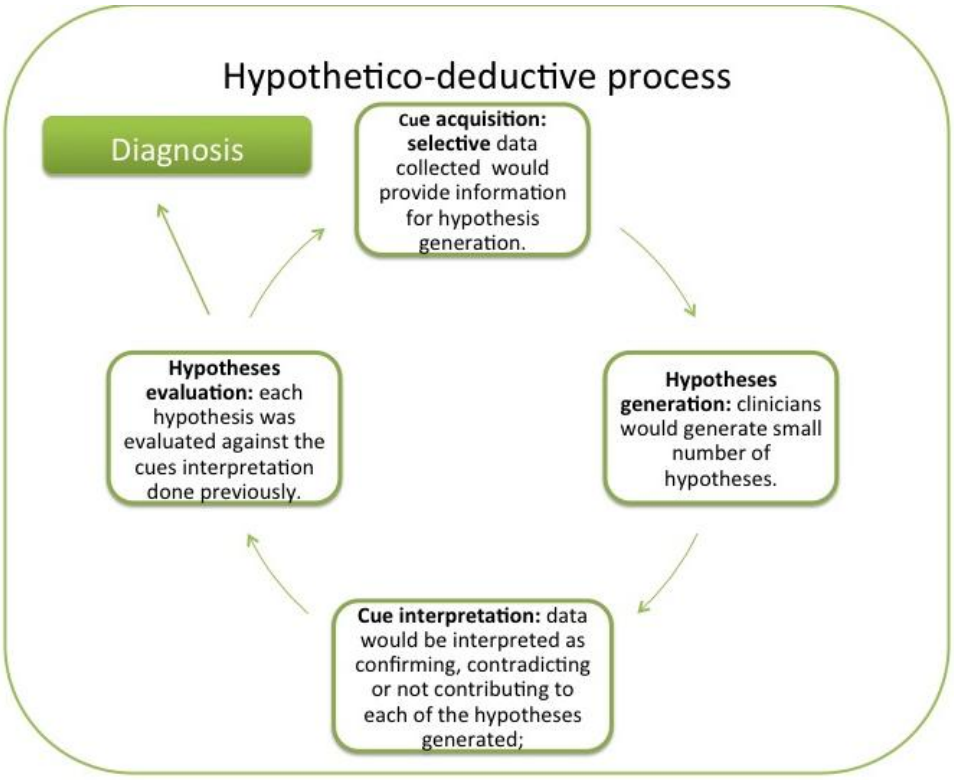
Hypothetico-deductive reasoning describes a process of human thought that is not exclusive of clinical reasoning (Holyoak & Morrison, 2005). However, in clinical reasoning research it acquires a specific meaning as to refer to a model of clinical reasoning first defined by Elstein & Schwartz and Barrows (Barrows & Feltovich,

6. Structure theories will be addressed in the next section: Clinical Reasoning Development.

1987; Elstein *et al*., 1978). This model suggests that clinicians would use early clinical data and previous acquired knowledge to generate hypotheses, and use those hypotheses to guide their further plan and test the adequacy of those hypotheses in an iterative process of "guide search" until the best possible solution was found (Elstein & Schwartz, 2002).

This reasoning process, shown in Figure 2. 2 comprehends four stages: "cues acquisition, hypothesis generation, cue interpretation, and hypothesis evaluation" (Elstein *et al*., 1978; Patel *et al.,* 2005 p. 7). Based on early clinical data, mainly from patient presentation, clinicians would generate a small set of hypotheses. These hypotheses would guide their selection of further data to be collected, which would be used to confirm or refute each hypothesis. If data would not confirm any of the hypotheses, the cycle would be repeated until a good enough hypothesis (diagnosis) was found.

Figure 2. 2- Hypothetico-deductive reasoning process (Elstein *et al*., 1978).



According to this model experts would differ from novices by starting with a more narrow number of best quality hypotheses (Patel *et al.,* 2005). This reasoning process would be a rational analytic process, in line to what is described by Dual-Processing Theory (DPT) as system 2 (Figure 2. 3). Describing a type of human reasoning that is a

conscious, controlled, high effort, analytic process based on clear rules (Evans, 2008). On this assumption the investigators believed this system to be accessible to clinicians conscious mind allowing for it to be studied using think aloud protocols (Ericsson, 2004).

However, Elstein and colleagues failed to find significant differences between experts (as judged by their peers) and non-experts in terms of their diagnostic (Ericsson, 2004; Norman, 2005; Patel *et al*., 2005). Also according to Elstein's models, more complete accounts of the rules used to make diagnosis would be provided by experts as those would be the basis of their cue interpretation phase. Research findings failed to show this difference, with experienced clinicians often providing very inaccurate and even more incomplete explanations for their decisions (Patel *et al*., 1986; Schuwirth, 2009).

Furthermore, other authors also found results that contradicted this explanation as the only way of reasoning used by experts to solve clinical cases by showing that experts often used similarities between new and old cases, drawing conclusions from those comparisons almost automatically (Norman *et al*., 1989).

Backwards Vs. Forward Reasoning

Backwards vs. Forward reasoning were first used in research into expertise in physics by comparing the strategies used by experts and novices solving mechanical problems (Larkin *et al*., 1980) and later lengthily explored in clinical reasoning research by Patel and Groen in 1986 (Patel *et al*., 1986; Norman *et al*., 1999). These authors' approach, as in the work from Elstein, was focused on applying generic models of reasoning and cognition theories to the study of clinical reasoning (Ericsson, 2007).

Data interpretation is central to the definition of these types of reasoning/clinical reasoning, with backwards reasoning in medicine describing the deductive process by which data is used to test, re-think, and evaluate initial hypotheses. As opposed to forward reasoning (in medicine), which describes the inductive process in which the hypothesis generation is driven by the data analysis of patients' results (Patel *et al*., 2005).

These authors, however, recognise that forward reasoning is not used in all circumstances, and suggested a reconciliation hypothesis (Patel *et al.,* 1990; Patel & Kaufman, 2000). They suggested that in more difficult cases, or cases out of the normal range expert reasoning "breaks down" shifting from forward reasoning into a

hypothetico-deductive or backwards reasoning model (Patel *et al.,* 1990; Patel & Kaufman, 2000). The *consensus model* was criticised by several as not providing enough explanation for the previous findings on clinical reasoning (Barrows *et al.,* 1982) and by their choice of cognitive psychology methods such as stimulated recall, protocol analysis in some studies. Such methods rely on small sample sizes and introduce bias associated with individuals' own ability to access their own thought process either simultaneously or after the events (Brooks *et al.,* 1991).

Non-analytic reasoning (or system 1)

This hypothesis was suggested by Norman and Barrows in 1997 (Norman & Brooks 1997) as a better explanation for the findings of some empirical studies of expertise (Norman *et al.,* 1985; Norman *et al.,* 1989). Namely that the fast and accurate way in which experts solved clinical cases did not seem to be totally explained by previous *analytic* explanations (Norman & Brooks, 1997). According to this view:

"Diagnostic thinking is based on the rapid and unconscious matching of the presenting problem to a similar, previously encountered, problem" (Brooks et al., 1991 p. 173).

This is based on the idea of specific-to-specific features (Brooks *et al.,* 1991), that is, by recognising specific features in a problem, the experts would then be able to retrieve past cases that would match those same features, and retrieve from memory all information related to that. This leads to a fast, not always conscious diagnosis (Norman *et al.,* 2007). These authors oppose the previous explanations, mainly the works by Elstein. In their view clinical reasoning is not an analytical process, but rather a faster, more intuitive and less conscious process (Norman *et al.,* 1989). Non-analytic clinical reasoning is based on system 1 of Dual Processing Theory (Leblanc *et al.,* 2001; Evans, 2008). System 1 *(*Figure 2. 3*)* is more automatic, rapid and less demanding of cognitive processing; it allows the clinician to deal with larger amounts of information simultaneously; it is holistic and more perceptual; it is domain specific, highly contextualised and has a pragmatic focus (Evans, 2008). This view has been supported by extensive research done by the McMaster group (Young *et al.,* 2007), and it is at the moment one of the most accepted explanations to expert clinical reasoning available (Ericsson, 2007).

Figure 2. 3 Characteristics of system 1 and system 2 according to Dual Processing Theory (DPT) (Evans, 2008 p. 247).

Characteristics of high order cognition systems according to dual-processing theories (Adapted from Evans,2008 p.257):

System 1	*System 2*
Evolutionarily: Old system, rationality, shared with animals, nonverbal, it is based on modular cognition.	*Evolutionarily:* Recent system, Individual rationality, Uniquely human, Linked with language, Fluid Intelligence.
Level of Consciousness: Unconscious (preconscious), implicit, automatic, low effort, automatic, rapid, high capacity, holistic and perceptual. It is default reasoning process.	*Level of Consciousness:* Conscious, Explicit , Controlled, High effort, Slow, Low capacity, Inhibitory, Analytic, reflective.
Functional characteristics: Associative, domain specific, contextualised, pragmatic, parallel, and stereotypical.	*Functional characteristics:* Rule based, Domain general, Abstract, Logical, Sequential , Egalitarian.
Individual differences: Universal, Independent of general intelligence or woking memory.	*Individual differences:* Heritable, Linked to general intelligence and Limited by working memory capacity.

Probabilistic, Bayesian Reasoning

Probabilistic and Bayesian reasoning models suggest that clinicians use an analytical approach to solving clinical cases (Oaksford, 2007). These models focus on how experts choose between hypothesis rather than on how those are formulated (Hahn and Oaksford, 2007). Bayes Theorem, for example, is a mathematical formula that can be applied to medicine to calculate a probability of diagnosis based on general information about disease prevalence, test sensitivity and specificity and test results of the patient to be diagnosed (Gigerenzer & Hoffrage 1995, 1999).

According to the Bayesian reasoning view, experts have very accurate *a priori* models of probability of presentation of particular diagnoses and conditional probabilities associated with each piece of evidence presented with the diagnosis. Based on this, and their assessment of the case, they are able to mentally compute the likelihood of the possible hypothesis under consideration (Oaksford, 2007).

A few problems have, however, been identified with this approach (Young *et al.,* 2011). First heuristics and biases are overlooked by this view. Studies in this area have shown that humans often make incorrect mental calculations (Evans *et al*., 2000). Research into the cognitive basis of medical error has shown that even experts lack the ability to correctly judge baseline probabilities from which their probabilistic reasoning would start (Berner 2011; Mamede *et al.,* 2010).

Hoffrage & Gigerenzer, for example, asked 48 clinicians to answer a simple question *"How many of those who test positive [in a mammography] actually have breast cancer?"* (Gigerenzer & Hoffrage, 1999 p. 425). Participants were given information about prevalence in population, women over 40, and test sensitivity and specificity in two different formats. When results were presented as probabilities only 8% of the clinicians were close to the correct Bayesian value, increasing to 46% when the results were presented as frequencies. Nonetheless the increase due to format presentation[7], the number of clinicians arriving to the correct prediction value was still less than half of the sample in this study (Gigerenzer & Hoffrage, 1999 p. 425). Secondly, some have suggested that small probabilities tend to be underestimated

---

7 Evans *et al*., 2000. Describes the ongoing debate on the impact of data presentation format ("frequency versus probability") on human ability to make estimations within the field human rationality and cognition.

and large probabilities tend to be overvalued not only in medicine but in other contexts as well (Elstein & Schwartz, 2002).

Finally a large amount of the empirical research that supports pattern recognition or non-analytic models of reasoning seems to refute the hypothesis of experts using highly elaborated mathematical probability systems (Young *et al.,* 2011).

Nevertheless the aforementioned studies with experts some have argued that teaching probabilistic reasoning and Bayesian reasoning rules can help students to make better judgments of disease probabilities and potentially reduce the likelihood of diagnostic errors at this novice level (Graber, 2009; Kurzenhäuser & Hoffrage, 2002; Sedlmeier & Gigerenzer, 2001).

A Consensus Model

Recently a more consensual view has been proposed, namely that experts use iterative reasoning to solve clinical problems, which means they are able to change from a non-analytic reasoning to a hypothetico-deductive reasoning when cases present a higher degree of difficulty (Croskerry, 2009a).

According to this perspective clinical reasoning would be routinely guided by pattern recognition, or non-analytic reasoning, however when facing problems of more difficulty experts *"go back"* to hypothetico-deductive type of reasoning mode (Norman *et al.,* 2009; Heemskerk *et al.,* 2008; Croskerry, 2009a). Therefore biomedical knowledge integration will be extremely important, as it will provide the fundamental basis of this second type of reasoning strategy (Boshuizen, 2003). Hypothetico-deductive reasoning would be accessible as a *"plan B"*, used in difficult cases or when a similar case is not available for retrieval from clinicians memory (Norman *et al.,* 2009; Croskerry, 2009b).

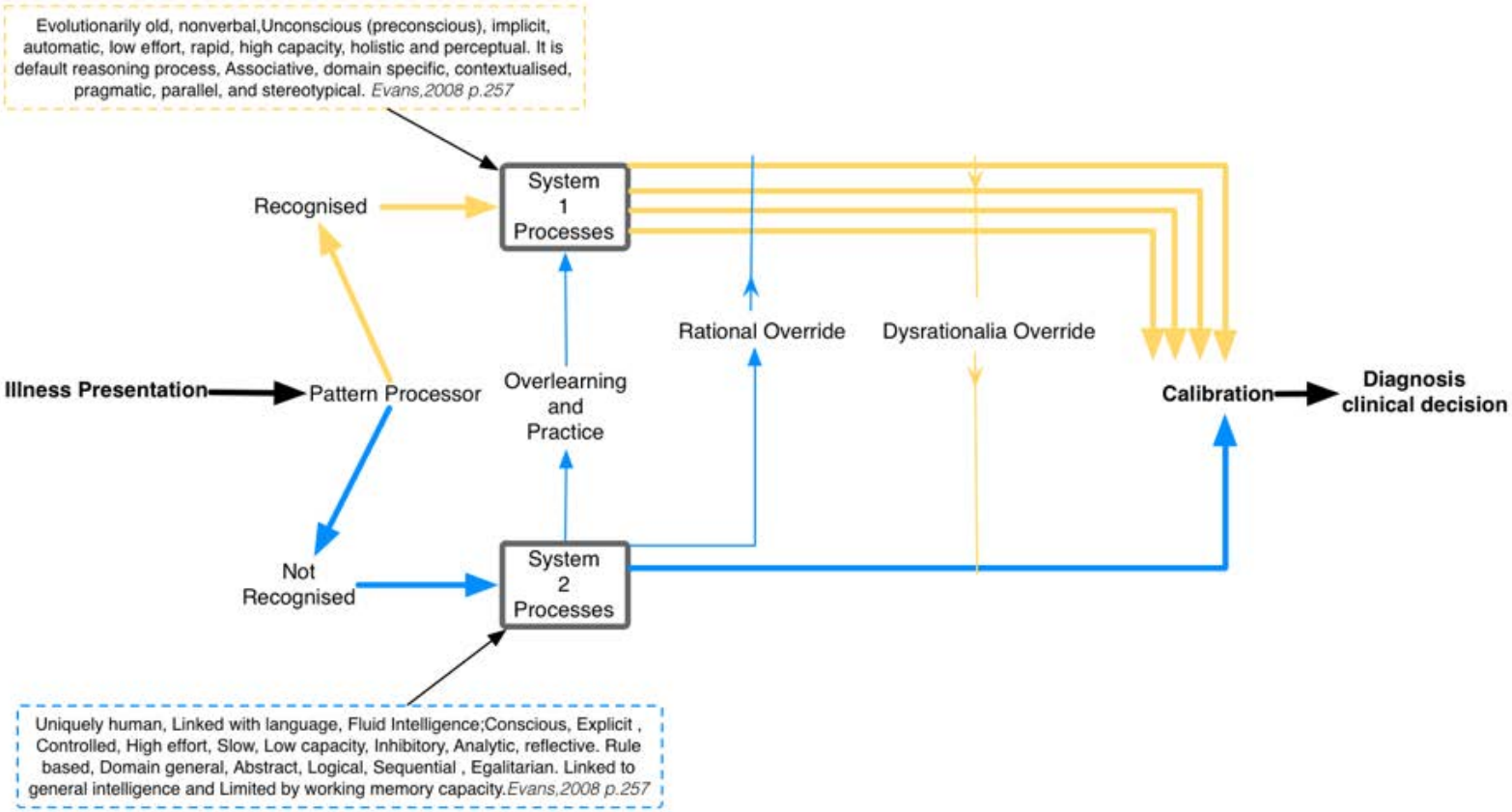
In this view novices should rely on the scientific approach using hypothetico-deductive reasoning to reach conclusions, if they attempt to use non-analytic reasoning they are likely to fail as they have not seen enough examples to compare new problems to (Norman *et al.,* 2009; Patel *et al.,* 2005). With practice, a database of cases would gradually form, and non-analytic reasoning would gradually start to become more capable of providing correct answers (Boshuizen, 2003; Norman *et al.,* 1999). The scheme below (Figure 2. 4) provides a schematic representation for this idea.

This view applies to clinical reasoning the principles of the Cognitive Continuum Theory (Cooksey, 2000; Offredy *et al*., 2008; Osman, 2004). This theory, as opposed

to the dual-processing theory (DPT) (Osman, 2004), argues that there is no dichotomy between system 1 and system 2 in human reasoning, but rather a continuum between *pure* intuition and analysis in which both elements interact assuming different weights according to contextual and situational factors (Offredy *et al*., 2008; Hammond 1996).

According to this "Universal" model, after a case is presented an internal "pattern processor" would allow the individual to very rapidly distinguish between recognised and not-recognised cases and determine if the expert would follow a non-analytic (system 1) or hypothetico-deductive (system 2) reasoning pathway. Importantly these systems are not mutually exclusive, they can interact, for example, a case feature may trigger a system 1 approach and a diagnosis to be generated almost automatically in the clinicians mind, but new information on the case may not fit that initial diagnosis requiring a thorough reassessment of the initial hypothesis (triggering system 2), the author refers to this process as *rational override*. Likewise there are abundant descriptions of situations where a carefully planned rational approach is overwritten by a more intuitive *"gut feeling"* on an individual situation or case, that is *dysrationalia override* (Croskerry, 2009a). Additionally system 2 may often monitor system 1 decisions. Analytic processes (system 2) such as reflection (Pakman, 2000; Schön, 1983) or metacognition (Quirk, 2006) are often used while solving clinical cases, and to monitor ones rapid and intuitive (system 1) decisions so that both systems can be updated and improved (Croskerry, 2009a).

Figure 2. 4: Schematic presentation of dual-processing model of clinical reasoning adapted from Croskerry, 2009 (p. 1024)



A.4. Clinical Reasoning: Development

Understanding the development of the mental process of combining known information (premises) and generating new information (conclusions) that represents the best possible solution for a patient's problem in a particular context and given time, is a jigsaw of variables and a very demanding task for researchers (Elstein, 2009).

The main focus of clinical reasoning research, as seen in the previous section, has been to understand *"how experts think"*, motivated by the premise that if one understood how an expert would reason when solving clinical cases then it would be possible to help novices to improve their own thinking processes. This is also related to the idea that perhaps clinical reasoning opportunistically develops with experience in the clinical setting (Kassirer, 2010) and possibly cannot be formally taught in medical school (Norman, 2005).

The extensive research done during the last forty decades on clinical reasoning, along with all the associated changes in medical education paradigms and medical practice have led to a shift in these ideas and a growth of interest in the development of clinical reasoning at the undergraduate level (Kassirer, 2010).

The teaching of clinical reasoning seems now much more of a responsibility of medical schools, being advocated by stakeholders (GMC, 2009) as an important part of the training of medical students (Woollard, 2006).

However, there is a remarkable scarcity of studies dedicated to the identification of the factors affecting the development of clinical reasoning in undergraduate medical education. The argument that active learning strategies, integrated subjects, case-based learning and even problem-based learning favour the development of clinical reasoning is merely presumptive with hardly any support or hard data (Albanese & Mitchell, 1993; Ark *et al.*, 2006; Bowen, 2006; Eva *et al.*, 1998; Patel *et al.*, 1994; Patel *et al.*, 1993; Schmidt, 1993).

Large sample developmental studies, based on longitudinal designs with frequent follow-ups, and multiple data collection instruments, as the ones traditionally used in developmental psychology and learning sciences (Jean Piaget's work being one of the greatest examples) would be very useful in attempting to understand this process and creating a suitable model and theory of clinical reasoning development (Bordage, 2007a).

Nevertheless despite the growing interest in clinical reasoning at the undergraduate level, to the present date, very few (for an exception see: Neufeld *et al.,* 1981) of those have been carried out, following medical students through their development into expert clinicians. Recent financial constraints on research departments and universities and outcome based research evaluations that value the number of publications, amongst other possible factors make longitudinal studies, that are long (in time) and expensive, less attractive to research institutions (Norman, 2002).

Studies which have been carried out aimed at understanding medical expertise and/or differences between experts and novices, have also provided explanations on how these skills develop through time and from experience (Norman 2005. Although expertise is a wider concept than clinical reasoning, the ability to effectively manage patient problems depends largely on clinical reasoning and therefore many expertise studies have drawn conclusions regarding its development (Mylopoulos & Regehr, 2007, Feldon, 2006).

From this vast body of research three main views on how clinical reasoning develops can be identified: Prototypes and Semantic qualifiers (Bordage & Lemieux, 1991); Encapsulation (Boshuizen *et al.,* 1992) and Illness scripts (Feltovich & Barrows, 1984) and Adaptive Memory and Pattern Recognition (Eva *et al.,* 2002).

Development 'models'

Prototypes and Semantic Qualifiers

Bordage's research used Elstein's work as an "early platform" (Bordage, 2007 p. 1117) but focused on knowledge organisation and knowledge structures (Bordage *et al*., 1990; Lemieux & Bordage, 1992; Bordage *et al*., 1990; Page *et al*., 1995). Based on previous works from Rosh (1975, 1976) and Mervis (1976) on categorisation, prototypes and memory, Bordage in 1984 proposed a prototype theory to medicine as an explanation for expert diagnostic skills (Bordage & Zacks, 1984). The basic assumption was that the difference between experts and non-experts would rely, not on a generic problem solving skill, but on the more efficient organisation of knowledge by experts, making retrieval processes less demanding cognitively, therefore faster and less error prone (Bordage, 1987; Bordage & Lemieux, 1991; Page *et al*., 1995).

Prototypes are defined as *"anchors"* for the retrieval of knowledge within a category. For example gastritis will be an anchor in the gastrointestinal disorders category

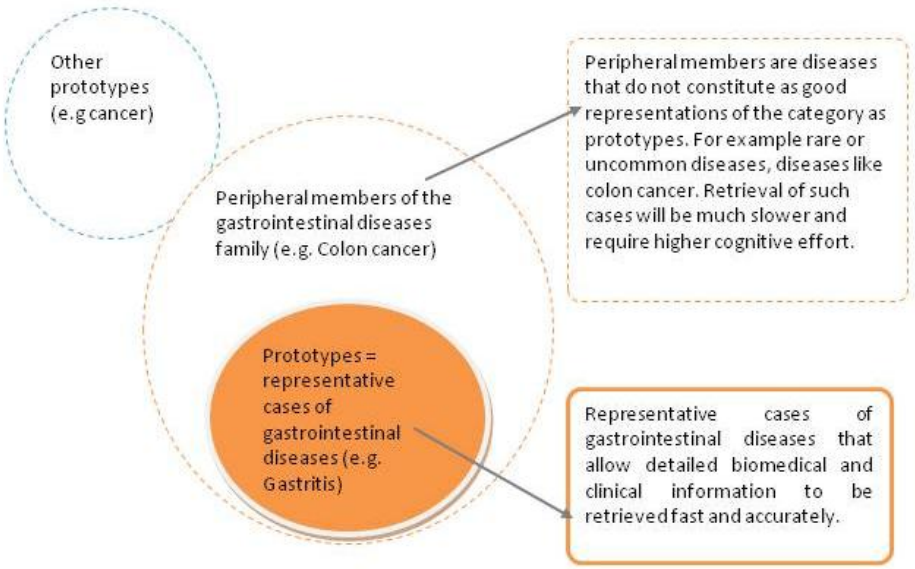
(Bordage & Zacks, 1984). When a patient presents with a gastrointestinal disorder, the node of knowledge related to gastritis would be retrieved (along with other anchors for the category) almost immediately and with great detail. If the presentation did not match these anchors, then more cognitive effort would be invested into retrieving other information about the category (e.g. gastrointestinal diseases) exploring other prototypes and peripheral members until a diagnosis was found as shown by **Error! Reference source not found.** (Lemieux & Bordage, 1992).

Bordage conducted four studies with medical students and family doctors (Bordage & Zacks, 1984) and two additional studies based on educational interventions designed to improve prototype formation with medical students alone (Bordage, 1987). From these studies the author concluded that there is sufficient support for prototype theory in medicine (Bordage *et al*., 1997; Bordage, 1987; Bordage & Zacks, 1984). However, these results were not clear enough to convince others as the relationship between prototypes and diagnostic ability was far from clear (Ericsson, 2007; Norman, 2005). Even Bordage, years later, recognised that this theory does not provide a suitable explanation for the connections between knowledge structures in memory. He proposed, then, that this organisation must be a product of semantic qualifiers, which can be viewed as dichotomous axis according to which concepts are classified, stored and integrated with previous networks. An example would be: *acute versus chronic* (Bordage, 1994) .

According to this proposal a higher number of semantic qualifiers would mean a more detailed and accurate classification, therefore leading to a higher degree of differentiation between concepts allowing faster and more accurate diagnosis (Bordage, 1994; Bordage, 2007b).

Experts would then use more semantic qualifiers then non-experts. Bordage observed, using thinking aloud protocols, (Bordage, 2007b) that successful diagnosticians used twice as many semantic qualifiers in their discourses as those who were not successful at the diagnostic task. This was also applicable to students (Bordage, 1994).

Figure 2. 5: Schematic representation of gastrointestinal diseases prototypes (adapted from Bordage (2007b).



The development of clinical reasoning would then take place according to two axis, the discourse and its semantic qualifiers, leading to four stages: 1st discourse would be reduced with limited semantic content, as students would not have much medical knowledge; 2nd would be a stage where there would be a dispersed or extended discourse with still limited semantic richness; new knowledge is assimilated but still mainly formed by disconnected facts; connections start to emerge; 3rd, elaborate discourse, knowledge database is now more substantial leading to a semantic richness but still insufficiently organised leading to a still extended or dispersed discourse; finally 4th stage, semantic richness and limited discourse; knowledge database has now been fully categorised according to semantic qualifiers, making it possible to easily identify relationships between concepts and build flexible clusters around those semantic qualifiers (Bordage, 1999).

This last stage would be similar to what was described by Schmidt and Boshuizen (Boshuizen *et al.,* 1992) as 'encapsulated knowledge'. The next section of this chapter will explain this view. The attempts of Bordage and colleagues to develop educational interventions that would promote this hypothesised clinical reasoning process did not prove successful. Inducing semantic representations did not seem to improve diagnostic accuracy (Bordage, 2007). Additionally some of the theoretical assumptions this framework was based on are questionable (Custers *et al.,* 1996b).

Categories are not always clearly decomposable into smaller opposing units, and finally it is difficult in this framework to explain the impact of the time on course of disease (Custers *et al.,* 1996b).

Even Bordage in a 2007 paper had moved away from his initial enthusiasm for this view, casting some doubts about its ability to provide an explanation for expertise and the development of clinical reasoning and advocating further research in the field (Bordage, 2007b).

"[Building upon exiting theories] has been and continues to be the case over the years in our work on prototypes and semantic qualifiers. Similarly, a single method is unlikely to provide satisfactory answers to all questions. Competing hypotheses and multiple, converging methods are more productive than single hypotheses and methods." (Bordage, 2007b p. 1119).

Encapsulation and Illness Scripts

Schmidt and Boshuizen (Boshuizen *et al.,* 1992) shared the focus of Bordage, (Bordage, 2007), Patel and Groen (Patel *et al.,* 1986) and Norman and Eva (Eva *et al*., 2002), by recognising that the structure of medical knowledge, its internal organisation and relationship with memory, are crucial to explaining medical expertise: it is not a generic "problem-solving trait". According to this author:

"Everyone, novice and expert alike, generates hypotheses in diagnostic reasoning and tests these hypotheses by gathering discriminatory information that either confirms or denies them. (...) in medicine at least but probably in the many other diagnostic professions, differences between experts and novices cannot be explained by differences in reasoning" (Boshuizen, 2003 p. 12).

Schmidt and Boshuizen (Boshuizen *et al.,* 1992) favoured Simon and Chase's (Chase & Simon, 1973) view on expertise development as opposed to Elstein's view (Ericsson & Ward, 2007). According to Simon and Chase expertise would be developed by acquiring complex patterns that would guide actions in similar situations.

Boshuizen suggested that expertise in medicine is acquired by three distinct phases in which knowledge organisation structures evolves. A first phase of *"knowledge acquisition, validation and integration"* (Boshuizen, 2003 p. 12), a second of *"knowledge encapsulation"* and a third phase involving the formation of *"illness scripts"* (Boshuizen, 2003).

The first phase, attributed by the authors to the early years of medical school, is characterised by lack organisation structures leading to dispersed storage of

knowledge in memory. This process is similar to one of the key elements described by Piaget's cognitive development theory as "assimilation, accommodation, equilibration" (Piaget, 1985). Here new information was acquired leading to the necessity of change in mental structures to be able to accommodate, or integrate new information, subsequently these changed/new structures allow the individual to make sense of the acquired information and enable relationships with previously stored information (Evans & Keenan, 2009; Jong, 2009). This is a highly demanding process in medicine as many domains of knowledge are involved simultaneously (Boshuizen *et al.,* 1995). The large body of medical knowledge, complex biomedical concepts, acquired skills and procedures are not yet organised in order to be easily available for activation and application to the reasoning process. Therefore the 'clinical reasoning' process at this stage is often very descriptive, lacking synthesis and often supported by notes and schemes (Boshuizen, 2003). Gradually students start to identify relationships between concepts, and scattered concepts increasingly form networks defined by the possible similarities and relationships between those concepts.

The next phase is knowledge encapsulation, referring to the automatic retrieval of concepts closely clustered together under a single designation. This occurs according to Boshuizen with the increasing use of previously formed networks of biomedical concepts. Clinical practice and exposure to real patients and the inherent situational demands of this context would provide an important element to the development of this process. By being exposed to real patients and concrete clinical situations these networks are frequently activated, tested and validated leading to the formation of simplified versions of the complex causal networks of biomedical concepts. Allowing the students to be able to make more faster connections between concepts without the cognitive demanding process of retrieving all the sequence of causal relationships involved; however, these, and the concepts involved can be retrieved from long-term memory if necessary (Boshuizen & Schmidt, 2000; Boshuizen *et al*., 1992).

An observable consequence of knowledge encapsulation is a change in language used by medical students (Boshuizen, 2003). As a result of encapsulation there would be an increase in the use of medical concepts (e.g. sepsis) to describe what previously would have been a long description of biomedical process (e.g. generalised infection cause by a severe response to bacteria). Allied with the use of concepts such as "micro-embolism", or "aortic-insufficiency" resulting from joining other concepts together (Boshuizen, 2003 p. 13).

Encapsulated knowledge provides a useful tool for the clinical reasoning process allowing students to reason much faster, with relatively less cognitive effort, but still following a step-by-step approach between networks of concepts.

A final phase, the expert level is characterised by illness-scripts formation (Schmidt *et al.,* 1990). This notion, introduced by Feltovich & Barrows in 1984, describes a pattern or sequence of action that is acquired from experience of similar situations. Illness-scripts contain an extensive amount of knowledge about the process of disease, like possible presentations, features, enabling conditions, possible situations misleading cues, consequences and actions required (Boshuizen & Schmidt, 2000; Charlin *et al.,* 2000). Additionally these also contain possible variations in the situations that activate them.

Due to the organisation of knowledge, when a common or previously seen situation that matches an acquired illness-script occurs its activation means that all the knowledge stored in it is retrieved, leading to faster and much less cognitively demanding reasoning. Often several scripts are activated simultaneously, allowing the expert to choose which one better suits the situation (Charlin *et al.,* 2000). This would makes experts much more likely to find a correct diagnosis than novices, as the amount of information they are able to activate and use simultaneously is much higher (Schmidt *et al.,* 1990).

Schmidt and Boshuizen and their team gathered considerable empirical support, from several studies in medicine (Schmidt *et al.,* 1990) and in other fields (Arts *et al.,* 2000).

However, in these studies they found that often the relationship between expertise and biomedical knowledge was not linear. Often final-year students or recently graduated doctors were able to provide more detailed and complete explanations for their reasoning, than even experts. This has been described as "intermediate effect" and interpreted by Boshuizen and his colleagues as evidence of knowledge encapsulation and scripts formation (Boshuizen *et al.,* 1992). According to this view final-year students and novice doctors would still be able to access the step-by-step networks, therefore providing more pathologically complete explanations, while experts would be using automatic illness-scripts, that are more focused on actions (Boshuizen, 2003).

This view has been criticised due the methods used (Schmidt *et al.,* 1990), the thinking-aloud protocols, stimulated recall and proposition analysis from which these conclusion are driven relies on the participants' ability to verbalise their thought

processes while they perform tasks. Thinking or talking aloud protocols do not hold enough power or reliability (Schmidt *et al.,* 1990, Eva & Lingard, 2008).

Thinking-aloud protocols are highly dependent on the participants' reasons for actions being accessible to their 'conscious mind', which is not the case for automatic, instinctive decisions (e.g. pattern recognition), therefore this method will only access those decision that have a strong analytic foundation. This means that the above explanation for the development of clinical reasoning would only be valid for to explain the development clinical reasoning using system 2 and not patter recognition (Norman, 2005).

Additionally, stimulated recall, has been proved to create false reports since a subject will report not what they were thinking but what, at the recall moment, they think is the most appropriate to say, creating false *post-factum* explanations (Norman, 2005).

Adaptive memory & Pattern Recognition

This view is presented by the works of the McMaster group headed by Norman. This group clearly opposes Elstein's work both in its focus and methodology.

According to the McMaster group studies expertise in medicine would be largely dependent on two factors, the ability to arrive at a possible diagnosis using the available information to confirm it (confirmatory) and the ability to use that information to discriminate between possible differential diagnoses (discriminatory) in order to select the most appropriate (Eva *et al.,* 2002). Confirmatory interpretations of data would be acquired first and only with expertise would the ability to use data in a discriminatory way arise.

"It could well be that the ability to seek confirmatory data arises fairly early, but that sensitivity to competing diagnoses is a higher level of skill, and therefore a more discriminating test of expertise." (Eva et al., 2002 p. 258).

Clinical experience will lead to storage of clinical cases in memory, novices and experts would then solve problems by comparing then to previous cases with new cases as a result of non-analytic reasoning (NAR) (Norman *et al.,* 2007), or system 1 as it is designated in cognitive psychology (Gobet, 2007). This process would explain differences between experts and novices, as experts would be matching the new case against a much more detailed and comprehensive database of previous cases (Schmidt *et al.,* 1990). Methodologically NAR is not accessible to memory, therefore stimulated recall or thinking or talking aloud protocols would not be able to reveal such a process (Ericsson, 2007).

The work of Norman and the McMaster group attempts to refute the knowledge encapsulation hypothesis as an explanation for knowledge organisation in expert diagnosticians (Custers *et al.,* 1996a). Expert memory performance will depend on processing done during the task and not by any previous development of categories (that is knowledge encapsulation) (Eva *et al.,* 2002). They have shown that there are differences in the retrieval of information depending on whether the clinical cases are similar to previously seen ones or not, meaning information retrieval is dependent on previous examples stored rather than on an activation of a diagnostic category previously formed (illness script) (Brooks *et al.,* 1991). The authors used a study involving students and clinicians, at three levels of expertise to support their view: medical students about to graduate (clinical clerks), residents, and internists (Eva *et al.,* 2002). In this study they presented the participants with a series of six clinical cases, containing symptoms and signs consistent with one pre-defined diagnosis. Participants were presented with a case and asked to generate diagnoses and return their suggestions to the researchers. They were then asked to list all the symptoms and signs that they could remember from that case (free recall), after which they were shown the diagnosis previously defined for the case and asked to list any reasons why this diagnosis was not in their diagnoses and if they had or not considered that diagnosis (cued recall). Finally they would be asked to select from a list of sixteen features, eight features that were present in the case history (recognition). This would be repeated for each case. Diagnostic accuracy and performance in these memory tasks (free recall, cued recall and recognition) were measured and analysed (Eva *et al.,* 2002).

Despite no significant differences in the percentage of participants making correct diagnoses in each of the groups (level of expertise), clinical clerks showed significantly better performance in the free recall task, and clerks and the residents performed significantly better then the internists on the recognition task. There were no significant differences registered in cued recall (Eva *et al.,* 2002). Although the focus of the study was the performances in the memory tasks the fact that no significant differences were found in diagnostic performance, with a slightly higher percentage of clerks (75.7%) then internists (75%) making the correct diagnosis, was overlooked. This result raises questions regarding the cases, and the selection of participants. It may be that the level of difficulty in the cases was favouring the novices and just being to simple for experts, therefore requiring less cognitive effort, which would be associated with less effective memory retention.

Nevertheless the results in memory tasks seem to be consistent with the intermediate effect described by Schmidt and Boshuizen (Schmidt *et al.,* 1990). However these authors argue that previous explanation for this effect of put forward by Schmidt and Boshuizen was insufficient to explain their results. The encapsulation hypothesis would postulate that experts remember the cases as well as novices, but because the information is organised in a more effective way around clusters, which leads experts to pay attention to only a few very significant features of the case and then retrieving from memory related clusters of information based on previous knowledge of similar cases. Eva and colleagues purposefully introduced in the recognition task false case features to test for this effect. These features would be consistent with the cases, but would not have been present in the case history. For example in a case of gastritis abdominal pain could be absent from patient's symptoms, but would be introduced in the recognition phase as a false feature. According to the encapsulation hypothesis experts would be more likely then novices to think these false features were part of the case, because these would be in the cluster of information retrieved in response to the case (Eva *et al.,* 2002). Their results counter this hypothesis, there was no difference in the number of "false alarms" between experts and novices, but internists performed worse (significantly) in the number of case correct features they recognised (Eva *et al.,* 2002).

This was also supported by the free recall results showing that clerks had a better memory of the cases' features, remembering better all types of information including key features of the case. Opposing the view that experts attend only to key pieces of information while novices gather all types of data without attending to their importance for the case (Eva *et al.,* 2002).

On the basis of these results the authors concluded that there was no evidence of knowledge encapsulation of cases by experts. Rather experts seem to process cases holistically as had been suggested before by Barrows (Barrows *et al.,* 1982). Additionally they seem to find no evidence to refute the idea that "*differences in performance on memory tasks result from changes in the adaptive nature of feature memorisation.*" (Eva *et al*., 2002 p. 261), which supports a view previously presented by Norman (Norman *et al*., 1989).

This view was based on the memory theory of Chase and Simon and on the Transfer Appropriate Processing (TAP)[8] model (Simon & Chase, 1973). According to this view

8. This model was first proposed by Morris, Bransford, and Franks (1977). TAP suggests that levels of

their explanation for experts would having better memory of the cases is because they see gathering examples and storing template cases in their memory as useful to improve their practice. While a novice would be trying to memorise all details about a case, the experts would try to identify the key aspects that could be transferred from that case to other cases and retain only those in their memory (Norman *et al*., 1989).

These conclusions support the view that experts use non-analytical reasoning (NAR) or pattern recognition to solve clinical problems (Norman & Brooks, 1997), a process not accessible to novices. Clinical reasoning would develop with the transition from a rational knowledge base, acquired via educational experiences, to a non-analytic ability to recognise and handle familiar situations (van der Vleuten & Newble, 1995). According to this view, novices have to rely on the scientific hypothetico-deductive reasoning approach, to reach conclusions, and when using non-analytic reasoning they are likely to fail as they have not seen enough examples to compare to new problems (Norman & Brooks, 1997).

Impact of clinical practice: Deliberate and reflective practice

"Learning strategies are like all other skills. They can be practised and it is very helpful when learners have some meta-cognitive understanding of their usefulness and the conditions for their application."(Boshuizen, 2003).

Students' exposure to clinical cases in the clinical context is very important to the development of clinical reasoning (Ajjawi & Higgs, 2008; Eva, 2005; Boshuizen, 2003; Bowen, 2006). In this section we describe jointly the ideas presented by Ericsson (Ericsson *et al.,* 1993) on the application of deliberate practice and reflective practice

---

processing can be manipulated depending on the task, its goal and purposes. Quality and durability of memory will be dependent on the material to be memorised and the test task designed, in other words precise differentiation of the acquisition and test stimuli. Superficial processing does not by itself ensure durability of information, as it is possible to retain information processed superficially for long periods of time in memory.

Blaxton's (1989) studies also show that performance in data-driven and conceptually driven tasks were not associated and that could be better explained by TAP than by the memory organisation approach. Graf, P. & Ryan's (1990) studies on TAP and memory also concluded that familiarity and integration fosters retrieval from memory because cues can trigger reintegration with previous knowledge and trigger recognition decisions.

Bransford, J. & Frank, J., 1977. Levels of processing versus transfer appropriate processing*. Journal of verbal

Blaxton, T.A., 1989. Investigating Dissociations Among Memory Measures : Support for a Transfer-Appropriate Processing Framework. Cognition, 15(4).

Graf, P. & Ryan, L., 1990. Transfer-Appropriate Processing for Implicit and Explicit Memory. Cognition, 16(6), pp. 978-992.

(Mamede & Schmidt, 2004) as these provide explanations for the processes occurring during the contact with clinical practice that lead to the development of expertise in clinical reasoning.

Ericsson studied expertise and experts in many contexts, presenting a theory of expertise that is not specific to medicine. In Ericsson's view, as in the early expertise studies of Elstein (Elstein *et al.,* 1978) and more directly in Mylopoulos & Regehr (Mylopoulos & Regehr, 2011). Experts are not only the ones who have been repeatedly engaged in an activity, they are the ones who demonstrate superior performance in that same activity (Mylopoulos & Regehr, 2011). One of the most common examples is in sport; most of us engage in some sport but only a few can see their names up for gold medals in the Olympic games: those are the experts. Identifying who the real experts are in medicine is highly complex, as Elstein (Elstein, 2009) concluded, due to the lack of objective feedback on medical "performance" over time. Additionally, peer assessment of expertise does not seem to be an ideal way as factors such as hierarchical position, age or post seem to have an impact on whether or not a doctor thinks a colleague is an expert (Bordage, 2007).

Ericsson's general theory of expertise, and its application, to medicine is formed around the concept of deliberate practice. This concept can be summarised by the idea that practice alone is not enough for expertise, it needs to happen under purposely designed conditions, meaning practice has to be seen as a means to an end, that is, achieving a better performance. This not only means that practice has to be intentional and not opportunistic, it has to be planned, evaluated and redesigned based on identified learning needs and it has to represent a challenge (Ericsson, 2006). Another important consideration of this view is that expertise has to be proactively maintained, while experience simple accumulates over years. Using the aforementioned example, if a gold medal athlete stops practicing they are unlikely to win another gold medal in the future. Ericsson argues that the same happens across domains and that expert performance will only be maintained with the appropriate intention, effort and planned actions. Experts in the medical domain achieve and maintain their superior performance by exposing themselves intentionally and frequently to challenging tasks and that is engaging in deliberate practice (Ericsson & Towne, 2010).

Reflective practice was first presented by Schön in 1983 (Schon, 1983). Reflective practice suggests that skilled practitioners, or experts, use reflections upon practice as a way to assess and revise existing models of action in order to develop more

effective action strategies (Osterman, 1990). One can argue that some similarities can be found between Ericsson's notion of deliberate practice (Ericsson, 2004) and the action planning component of Kolb's theory (Kolb, 1984; Kolb *et al.,* 2000; Smith, 2001; Baker *et al.,* 2002; Kolb & Kolb, 2005). Expertise in professional domains develops by intentional and frequent practice, and a simultaneous reflection driven by those actions. This reflection will lead to changes in future plans and actions, and by engaging in this continuous process, experts will achieve and maintain their superior performance (Mamede & Schmidt, 2004; Feltovich *et al.,* 2006; Ericsson, 2004).

"Reflective practice is conceived as the ability of doctors to think critically about their own reasoning and decisions" (Mamede *et al.,* 2008 p. 468).

By this definition it is possible to observe the close connection between reflection and clinical reasoning with the assumption that reflection would be associated with an improvement in clinical reasoning (Epstein, 1999).

This is of course a highly demanding task in the medical field, where decisions have to be made at a very fast pace, and based on a high degree of uncertainty (Kuipers *et al*., 1988). Therefore Mamede and Schmidt propose a five-factor model of reflective practice in medicine, based on answers of a group of 202 Brazilian primary care doctors. The five factors proposed are: "deliberate induction; deliberate deduction; testing and synthesising; openness for reflection, and meta-reasoning" (Mamede & Schmidt, 2004 p. 1302). Deliberate induction, means the intentional search for further alternatives in addition to the generated hypotheses when solving a complex problem. Deliberate deduction would be the exploration of possible scenarios, the consequences of those alternative explanations that can be tested against new data. Testing and synthesising is a process of testing previously made predictions against data collected from the case and summarising them into only a few key elements. Openness towards reflection, means that there must be an intentional effort to engage in reflective strategies to solve the case (Mamede & Schmidt, 2004).

Meta-reasoning, would describe the ability to reflect not only on the actions but also to assess and judge one's own reasoning processes, evaluating your assumptions and conclusions (Mamede & Schmidt, 2004; Mamede, Schmidt & Penaforte, 2008b). Mamede & Schmidt argue this would provide a useful framework to identify, measure and teach students to become reflective practitioners (Mamede & Schmidt, 2004 p. 1302).

Importantly these reflective or deliberated processes are conscious processes that would take place within the system 2, as defined by Dual Processing Theories, as they would require a level of consciousness, analysis and planned thinking only possible within this system (Evans, 2008; Young *et al*., 2007). Due to their increasingly recognised importance, Stanovich argues that there are actually two subsystems at play within the system 2. One responsible for a 'rational' processing engaging with the 'algorithmic mind', intellectual abilities and traditional analytic strategies, the other a 'reflective mind' with would involve beliefs, attitudes and 'thinking dispositions' (Stanovich, 2010a, p. 25). This 'reflective mind' would operate at a higher level and have the ability to resolve conflicts between system 1 and system 2 (Stanovich, 2010b). Both subsystems are considered within Croskerry's model, not as separated instances but as functional elements operating within system 2 mainly involved in 'rational override' of system 1 process when these prove not to be adequate and in contributing towards updating and facilitating the integration of new experiences within the individual cognitive processing systems (overlearning and practice) (Croskerry, 2009a).

A study with postgraduate medical students (interns) from Mamede & Schmidt also supports for this idea. These authors demonstrated that their aforementioned framework allows interns to improve their ability to deal with difficult cases in uncertain contexts. Additionally also concluded that reflective reasoning seems to be as effective as non-analytic reasoning when dealing with routine cases, refuting the idea that NAR is a more effective way to deal with these cases (Norman *et al.,* 2007).

Additionally reflective and deliberate practice theories make explicit the fact that experience alone is not enough to guarantee continuous development in the performance of clinical reasoning. To some extent this seems to contradict, once more, the views on NAR (Mamede & Schmidt, 2004), because in the latter view experience alone would lead to exposure to a bigger pool of cases and therefore an increase in the number of illness scripts, or "case templates" used to match and solve new cases. While the first mentioned theories, suggest that individual intentions, dispositions, structure and meaning of the experience influence this learning process and can be determinate on acquisition of expertise (Kolb, 1984; Kolb *et al.,* 2000; Smith, 2001; Baker *et al.,* 2002; Kolb & Kolb, 2005; Ericsson, 2004; Schon, 1983).

A.5. Clinical Reasoning: Assessment

As previously described in this chapter (see section *B. Clinical reasoning a definition*) clinical reasoning, maybe due to its complexity, has been described in a slightly different way by different authors with no single agreed definition possibly reflecting the assessment methods used to measure it. Additionally clinical reasoning, as with any other type of high order cognitive process, is not available to direct observation, so it can only be assessed indirectly through the assessment of its products or via individual verbalisation (Patel *et al.,* 2005). Finally, the nature of Medical Education as a multidisciplinary research field populated by researchers from several different backgrounds using different scientific paradigms, accepting a range of possible solutions to the same problem, means it is a non-paradigmatic endeavour (Norman *et al.,* 2009).

These three factors help us understand why, so far, no single instrument or method to assess clinical reasoning has been accepted, but instead multiple instruments and methods have been used. This section will cover these methods, by dividing them into three main categories according to their nature, purpose and aim.

A first category concerns the experimental and research methods from behavioural and cognitive sciences that have been used to study clinical reasoning. These methods are widely used as ways to provide insight into cognitive processes and are usually linked to experimental design sets. Methods such as thinking/talking aloud protocols, stimulated recall or protocol analysis, are among the most frequently used (Lundgrén-Laine & Salanterä, 2010; Ericsson, 2008) to study clinical reasoning. Here the aim was more to understand the process of clinical reasoning, for example by comparing novices and experts strategies.

A second category is the instruments that were purposefully developed, during the last forty years, with the specific aim of assessing clinical reasoning. These were developed on the premise that by measuring an external outcome it was possible to gain understanding of the internal processes behind reasoning. These also represented a need from authors and their institutions for introducing clinical reasoning into their assessment agendas in order to make sure that undergraduates or postgraduates students were mastering this skill at their expected level.

Finally some have advocated that new technologies, such as high-fidelity simulations, virtual learning patients and virtual reality can contribute towards clinical reasoning

assessment. The opportunities and challenges these bring upon clinical reasoning assessments will be briefly summarised at the end of this chapter.

A.5.1 Experimental/Research Methods

Generally these are referred to as verbal protocols. These include three major types of data collection methods: stimulated recall, thinking and talking aloud protocols and protocol analysis.

Thinking aloud, protocol analysis and stimulated recall are verbal protocols methodologies that have been extensively used in cognitive psychology to gain access to individuals mental processes, e.g. thinking and reasoning (Ericsson, 2008). While stimulated recall happens after the action, both protocol analysis and thinking aloud protocols are concurrent with the performance of the task (Lundgrén-Laine & Salanterä, 2010). These last two methodologies are based on the assumption that with proper training, cues from the researcher in carefully designed situations, individuals are able to produce verbalisations that will not affect their performance. These will then allow researchers access to the individuals' short-term and working memory (Ericsson & Simons, 1993), allowing conclusions to be drawn from such verbal protocols (Ericsson, 2006; Ericsson, 2008). For these reasons, these methodologies have been frequently used in studying expertise in medicine and clinical reasoning research (Lundgrén-Laine & Salanterä, 2010). They have been used alone or combined with scenarios of patients or clinical cases, as part of more extensive research programmes using multi-method approaches (Ericsson & Simons, 1993) and/or combined with educational activities such as high fidelity simulation (Ericsson, 2008). All the major research programmes have used this type of methodology in their studies (e.g. Norman, Ericsson, Boshuizen, Schmidt, Patel). However an important critique must be noted, all of these research methods rely on the individual ability to verbalise thought processes. On one hand this is always limited by the participants verbalisation abilities, on the other it is limited by the availability of those thought processes to conscious inspection and recall.

Finally, it is important to note that all protocols are research methodologies, they do not intend to measure performance, or whether decisions taken at a certain moment of the task are correct (Ericsson, 2008). Their aim is to contribute to understanding the mental processes leading to performance, in order words the processes and strategies used by individuals (Ericsson, 2008).

Stimulated Recall

Stimulated recall, also termed a retrospective thinking aloud protocol consists in asking a participant to perform a task, record the performance and then asking the participant to explicitly verbalise to the researcher what he/she was thinking at each moment of the task (Ericsson, 2006). A few studies used this method; however, as research developed several critics to this technique were highlighted (Ericsson, 2006). One of its main critiques was due to the fact that participant reports would be highly inaccurate as they would focus only on the end points once the task was finished, and had difficulty in reporting concurrent views that led to their conclusions (Ericsson, 2006). Additionally it was also often found that with this method participants would reconstruct their own thought processes to match their own views on how they should be thinking, building after-the-fact reconstructions of their own performance (Ericsson, 2006). Controlling for these *post-factum* generated explanations is virtually impossible therefore this method was gradually replaced by other methods such as thinking aloud protocols or protocol analysis (see below), both in cognitive psychology (Wiley & Jee, 2011) and clinical reasoning research (Norman, 2005).

Thinking Aloud

Thinking aloud protocols consist of asking a participant to verbalise their thought processes while simultaneously performing a specific carefully designed task. This method proved to be a much more valid and reliable way to access mental processes (Ericsson, 2010) because, on the one hand the concurrent nature of the verbalisation allows a collection of data not only on the end products but also on the processes leading to those. In addition all the interpretation is done by the researcher and not by the individual eliminating the possibility of retrospective self-reconstructions of the thought processes (Stefano *et al*., 2010).

This protocol allows the collection of extensive and rich data on the processes involved in tasks, but in order to code and analyse these data researchers have to be experienced and trained, otherwise the method can be compromised (Lundgrén-Laine & Salanterä, 2010). Equally participants need to be trained to verbalise, otherwise the verbalisation process may impact negatively on performance

(Lundgrén-Laine & Salanterä, 2010). Warm-up tasks and adequate level of cues provided by the research are suggested as way to deal with the cognitive load aspects and improve verbalisation (Hayes, 1986).

Protocol Analysis

Protocol analysis was used by Watson (1920) and Dunker (1945) as a new method to analyse verbal reports (Ericsson, 2006). In protocol analysis participants are asked to follow a thinking aloud protocol involving concurrent verbalisation of their thoughts while performing a task. The data collected from this verbal report is then compared to a task analysis conducted previously that contains possible alternatives of action for that task (Ericsson & Simon, 1984). By adding the task analysis element to the analysis of verbal outputs produced during the thinking aloud process protocol analysis increases the validity of the coding of the verbal reports and provides a useful framework from which conclusions regarding the strategies and process of thought can be drawn (Ericsson & Simon, 1984). Additionally task analysis has a large input from the literature on the subject, being theory-driven, that allows models and theories to be tested against participants verbalisations (Ericsson & Simon, 1984). Thinking aloud and protocol analysis have been extensively used in clinical reasoning (Ericsson, 2002) and are today:

"central to the design of surveys and interviews (Bradburn and Schwarz, 1996) *and user testing of computer software"* (Ericsson, 2002 p. 4).

One important characteristic of both thinking aloud and protocol analysis is that the verbal reports happen in experimental settings, and participants are observed and recorded on predefined tasks usually in controlled environments (Ericsson, 2007). This provides validity to the data collection and the analysis of the data. However, this fact also limits the sample of verbalisations, as time and the recurrence of tasks is limited and might raise a few questions regarding the ecological validly of observations (Ericsson, 2007).

A.5.2 Clinical Reasoning Assessments Designed for Purpose

An overview of the assessment tests and instruments developed specifically with the purpose of assessing clinical reasoning over the last 30 years of clinical reasoning

research is presented in the table below. Contrary to the methodologies described before, these instruments were designed to assess performance in solving clinical cases or in particular steps leading to this solution.

The following Table 2. 1 presents a summary of the main instruments that have been designed with the specific purpose of assessing clinical reasoning.

Figure 2. 6: Clinical reasoning assessment timeline

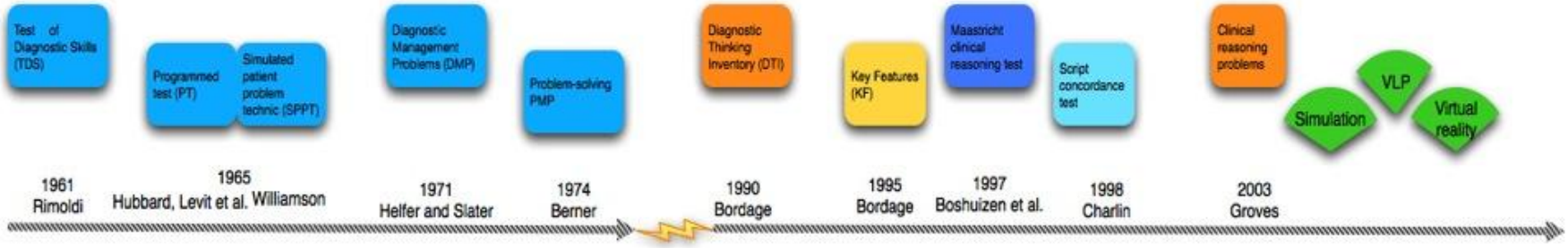


Table 2. 1: Description of clinical reasoning assessments

	Authors	Year	Description and principles	Strengths	Critiques	Studies
Patient Management Problems (PMP) (Versions of PMP: Sequential Management Problems (SMP), Branched Management Problems (BMP)	Rimoldi. Later version by McGuire and Babbott (1967), Helfer and Slater (1971)	1955 1967 1971	Clinical reasoning is seen as generic quality, a person's attribute reflecting the ability to solve problems in clinical domain; Clinical problem-solving reasoning is independent of knowledge and clinical skills; It can be measured by asking students to respond to problems standardised with objective scoring answers (Van der Vleuten and Newble1995); Consists of a sequence of choices made upon an initial patient description. The pathway though the problem is determined by the choices of the student. The choices can lead to the death of the patient, "patient expires. End of the problem" (=feedback to the student) Step 1: initial description of a patient; Step 2: Students chose what kind of data to gather, what tests to order, what therapy to prescribe from a large array of possibilities;Step3: Students get immediate feedback on their options; Step 4: they proceed to next steps of the problem (Berner, *et al.,* 1974).	Objective, easy to administer, simulated real problem-solving (Berner, *et al.,* 1974). This method can be at the same moment an assessment and an educational instrument. (Berner, *et al.,* 1974); Addresses all the components of the clinical cycle.	The basic assumptions were proved to be incorrect, as no generic problem solving skills independent of knowledge was found in medicine. (Elstein *et al*., 1978); Content specificity criticism: "no further investigation are required to demonstrate convincingly that one cannot generalize about so-called problem-	This method is a form of the Tab-Test Technique developed by the U.S. army. Introduced into medicine by the NBME and further developed by Christine McGuire at the Illinois College of Medicine; University of Illinois College of Medicine; Berner, *et al.,* 1974; Errors are cumulative (Berner, 1974); Because it is based on a decision tree schema, it

					solving skills from an individual performance "; High intra-individual variance on PMP (McGuire 1985)	provides additional guidance to the reasoning and it leads students to use a deductive reasoning strategy. Students do not have to state clearly what is the diagnosis, they just have to make correct decisions about taking account of the case presented and the feedback they get Van der Vleuten and Newble1995).
Test of Diagnostic Skills	Rimoldi	1961	"The purpose of this test is to estimate how a medical student proceeds when diagnosing a clinical case.(..) reach a diagnosis by asking those questions that he considers pertinent."(Rimoldi, 1961). Consists in two steps: Step 1 : Participant has some preliminary information about a case. Step 2 : Participant requests additional	It is more authentic, simulates all the clinical cycle; Enables one to see what information	Expert doctors, can get to the diagnosis using less information than novices or students, and by asking fewer	No major studies, apart from the authors.

			information (if necessary).Group of cards contained in flat pockets which particularly overlap and are evenly arrange on a display folder. On the top edge of each card a question that the examiner may ask is indicated (Rimoldi, 1961).	students use to diagnose a clinical case . It was designed for undergraduate medical students.	questions, but usually more important or decisive ones (van der Vleuten and Newble, 1995). Scoring number of questions implies that all of the subjects were using a deductive reasoning strategy. Additionally the basic assumptions were proved to be incorrect, as no generic problem solving skills independent of knowledge was found in medicine. (Elstein *et al*., 1978).	

Programmed Tests	Hubbard, Levit, Schumcher & Schnabel	1965	Reading a test booklet of clinical information about a patient, making decisions about the diagnosis and selecting the investigations they would perform to confirm it. The aim was to measure student's clinical judgment. The results were given. Subjects had to present further management decisions, then they were score by the number of correct decisions they made. (Berner, 1974).	It is more authentic, simulates t all the clinical cycle and it can be used with large samples.	It didn't show evidence of being an effective method to assess clinical reasoning or judgment. Instead it seemed to be measuring knowledge . Complex to develop and score because of the "correct decision" factor . Assumes there is a generic problem solving skills independent of knowledge, that was found not be correct in medicine. (Elstein *et al*., 1978).	Hubbard, Levitt, Schumcher & Schnabel, 1965; National Board of Medical Examiners. U.S.A.
Simulated patient-	Williamson	1965	This technique is based on the PMP's. It is a development of this test designed to assess the	Provides opportunity for		No major studies, apart from the

problem technique			product and the process of clinical competence. Step 1: Simulated patient test booklet. Step 2: choose whether to: Obtain a brief history/Perform physical examination/Order laboratory studies/Start therapy. then the subject has to do a sequence of decisions in order to gather as much information as necessary. This is a cumulative process. Like a decision tree the choices will have implications on the information presented. The decisions are compared with a criteria established by specialists, 2 indices are established: Efficiency Index (%) of "helpful solutions" and the Proficiency Index (%) of agreement with the criterion group in selecting beneficial and avoiding harmful interventions. CI (competence index)= {(PI x EI/100) + PI} 2 (William, 1967).	the physician to make clinical decisions like those he would make in his practice. The answer sheet provides feedback and simultaneously provides a record of the physician's management skills (van der Vleuten and Newble, 1995).		authors
Case Study Problems	Fleisher	1972	1st page: instructions, method of evaluation and the normal ranges lab values for 10 tests; 2nd page: Vertical columns presents information about 10 patients (Age, sex., race, temperature) the results of 10 (previously presented in instructions) lab tests. A chest and long-bone X-ray are exhibited on a viewing box to which	Originally developed to measure visual discrimination and skills in the interpretation and application of	Extremely focused on lab tests and interpretation. No reference to chief complaints of the patient or to information	No major studies, apart from the authors.

			students may choose to go for closer examination. 3rd page: A list of 100 numbered items (specific diagnoses, pathophysiological states, additional laboratory tests, diagnostic evaluations and consultations) arranged alphabetically. The students write the numbers of one or more of the 100 items appearing on page 3 at the bottom of the vertical column of the patient in the page 2, if that particular item is thought to be very pertinent to the case in consideration. The scores were based in experts' opinions about the pertinence of the 100 items of page 3 from +5 to +1 or 0 for no relevance.	lab studies. According to the authors it appears capable of evaluating cognitive capacities beyond those of recall or discrimination. Feasibility: not expensive to apply: easy to answer; It appears to be a reasonably reliable instrument.	presented for the patient. It is mainly driven by a conception of clinical reasoning as a) problem solving b) deductive reasoning. It is not as easy to score as a MCQ.	
Diagnostic Thinking Inventory	Bordage	1990	Self-reported scale that measures the degrees of flexibility in thinking and the degree of knowledge structured in memory.	Largely used, easy to mark and use. It Is independent of knowledge. It has gathered substantial support (van der	Self reported behaviour Eriksson, Smith, 1993), 'reasoning' is hypothesised to occur with sufficient	Goss *et al.*, 2011.; McAleer, 2008.; Beullens, Struyf & Van Damme, 2006; Groves, Rourke & Alexander, 2003; Groves *et al.,* 2002;

				Vleuten and Newble, 1995).	automaticity to make it take place without conscious awareness (Eva *et al.*, 2005). Does not measure actual behaviour or behavioural outcomes (van der Vleuten and Newble, 1995). Does not prove to be a reliable way to measure clinical reasoning, and to distinguish different types of clinical reasoning, or strategies used (van der Vleuten and Newble1995).	Groves, 2005; Sobral, 1995.
Key Features	Bordage	1992	According to this approach handling a clinical case depends on a few critical elements. So the test involving clinical competence or reasoning	It is a feasible approach to the complex reality of	Doctors were found to collect different amounts	Cook & Triola, 2009; Steinert *et al.*, 2006; Raupach *et al.*, 2009;

			should focus on those crucial elements. The problem is a clinical case scenario followed by questions that focus only on critical steps. The examinees can either write in their responses or select from a list of options (van der Vleuten and Newble1995)	clinical tasks (van der Vleuten and Newble 1995) Difficult to define the optimum data-collection strategy and the pathway used by the more experienced doctors.	of information and use different paths though the problems; Difficult to score complex cases due to inter-subjectivity among experienced doctors. Problems of validity and reliability - numerous weighting schemes were developed; Component scores were defined but small differences were found among them.	Cavaco *et al.*, 2011; Arts, , Gijselaers & Segers, 2006; Page & Bordage, 1995; Bordage *et al.*, 1995. This was also used by the Medical Council of Canada.
Clinical Reasonin	Groves	2001	Clinical history and some data from the physical	Defensible as the best way to	It focuses mainly on clinical	Groves, Scott & Alexander (2002)

g Problems (CRP's)			examination are presented. Subjects have to present 2 possible diagnoses, listing the data used to get to such conclusions	evaluate diagnostic reasoning, but further studies are yet to be conducted to prove the psychometric power of this instrument.	investigation judgements.	
Script Concordance Test	Charlin *et al.*	1999 2000	SCT test is an assessment tool that measures the capacity to solve ill-defined problems, that is, reasoning in the context of uncertainty.	Authentic clinical case- vignette (does not contain all data required to provide the solution). Several options should be considered ; The response format allows for multiple cases to be used in the	It has gathered evidence as a way to assess ability to reason from clinical data, but not its ability to assess diagnostic performance as a whole, nor hypothesis generation. Also its value for assessing	Caire *et al.*, 2004; Cohen *et al.*, 2005.; Bland *et al.*, 2005; Lebeau & Pagonis, 2006; Smith, 2008; Lambert *et al.*, 2009, Boulouffe, *et al.*,, 2010; Goulet *et al.*, 2010; Lubarsky *et al.*, 2011; Zavaleta-HernÃ¡ndez *et al.*, 2011

				same test moment Good psychometric qualities) Reliability, validity of the test and the scoring process have been demonstrated	undergraduate students still lacks evidence. It has proved to be effective in the assessment of postgraduate or professional development. .Difficult to mark or standard set (Lubarsky *et al*., 2011)	
Comprehensive integrative puzzles (CIP)	Ber, R	2003	Extended matching crossword puzzle, with several clinical vignettes	Authors mentioned it has good reliability and validity; Includes all the steps of the clinical cycle.	Used by very few studies (one) with only face validity reported.	Manzar & Al-Khusaiby, 2004.

As it is possible to see from the table above many attempts were made to create an instrument, test or tools to assess clinical reasoning. The early instruments were based on adaptation of patient management problems (PMP), created by Rimoldi (Rimoldi, 1961).The PMP type of instruments would have at their core a long clinical case, usually covering all the steps of the clinical cycle from patient presentation to treatments and management (Rimoldi, 1961). Therefore these would contain only few cases, as it was believed (at the time) that clinical reasoning, or clinical problem solving was a generic skill similar to a personality "trait". Therefore the performance in one case would be a good predictor of performance in all cases (Rimoldi, 1961). With Elstein's studies in the late 70s and developments in cognitive psychology (Wiley & Jee, 2011) this proved not to be true (Norman, 2000). These findings led to the previous mentioned conclusions about the importance of clinical reasoning content, case and context specificities (see section *B. Clinical reasoning a definition* of present chapter) (Norman, 2000).

Therefore those instruments were dismissed in favour of instruments that, although not covering equally all the part of the clinical cycle, would allow an increase in the number of cases and situations presented at each assessment moment. These include the key features approach of Bordage (Bordage *et al.,* 1995; Page & Bordage, 1995) clinical reasoning problems by Grove (Groves *et al.,* 2002) and Script concordance test by Charlin (Charlin *et al.,* 2000).

Also of note is the Diagnostic Thinking Inventory presented by Bordage (Bordage *et al.,* 1990). This instrument is clearly distinct from all the others by not being a case-based instrument, also due to the fact that this is based on participants' self-report of clinical reasoning processes rather than measure of outcomes, the DTI it is independent of knowledge (Bordage *et al.,* 1990). This instrument classifies clinical reasoning strategies reported according to flexibility of thinking and medical knowledge structure, here experts are believed to have a greater thinking flexibility, that being able to adapt their thought strategies according to different situations and a more organised knowledge structure (Lemieux & Bordage, 1992). Because this is a measure of individual perception, many other individual variables (e.g. personality, confidence etc.) may impact on the final result. Additionally as it is not directly linked with a contextualised or specific task but rather a more generic situation this might lead the participants to make choices on the basis of a generic intention of action (Eva *et al.,* 2004). Finally, it requires the participants to have good meta-reasoning

skills (Eva *et al.,* 2004), which prove to be difficult to develop (Ericsson, 2007) and which requires the reasoning process to be consciously open to analysis which is not the case with non-analytic reasoning or system 1 (Norman & Brooks, 1997).

Finally one other instrument was described in literature the comprehensive integrative puzzles (CIP) (Ber, 2003). This test is based on the "extended matching crossword puzzle". According to its authors:

"The left-hand column contains diagnoses or brief clinical vignettes. To complete the cells of the grid the student is required to match, stepwise, the various 'disciplinary investigations' to the diagnoses or clinical vignettes. When the puzzle is completed each horizontal row reflects a coherent medical case. The completed horizontal rows reflect integrative ability (diagnostic thinking and clinical reasoning) and the vertical columns measure the student's proficiency in interpreting medical history data, physical examination findings, laboratory test results, ECG, imaging, special tests, pathology and pharmacology" (Ber, 2003 p. 171).

Although the instrument is almost nine year old, and the authors mentioned it has good reliability and validity, only one small study was found reporting its use (Manzar & Al-Khusaiby, 2004). This study describes an intervention and only face validity and student's interest are reported, therefore it is not possible to know whether it support or not the claims of validity of the CIP (Manzar & Al-Khusaiby, 2004). Without more research using this tool it is not possible to have a clear picture of its strengths and weaknesses.

From this review it is therefore possible to conclude that no single instrument presented in the literature can be considered a consensual tool to assess clinical reasoning, for all have strengths and weaknesses (Schuwirth & Van Der Vleuten, 2005; Schuwirth, 2009).

The studies that have looked into differences in students' clinical reasoning strategies have done so using experimental/observational methods, mainly verbal protocols (see previous section F1. Experimental/research methods), with small or medium size samples (Schuwirth, 2009). In the present research our aim was to compare the impact of three different curricula in the development of clinical reasoning at undergraduate level, not only by comparing simple outcomes of the reasoning process, but also by understanding the different strategies students were using to arrive at the outcomes. None of the instruments described in the literature seemed to be able to collect such data; therefore, a clinical reasoning test was designed for the purposes of the present research. This is a theory-driven test, designed based on

a carefully chosen combination of the major strengths of past and current instruments.

A.5.3 Clinical Reasoning Assessments: Other Methods

More recently some authors (Lane & Slavin, 2001; Kuipers *et al*., 2008) have suggested that high fidelity simulation, virtual learning patients and virtual reality can contribute towards assessing clinical reasoning. High fidelity simulations, designed purposefully for clinical reasoning, might become a useful tool for teaching, assessment and research on clinical reasoning (Leblanc *et al*., 2011). Likewise virtual learning patients (VLP), which consist of interactive cases presented online (Cook & Triola, 2009; Botezatu *et al.,* 2010) seem to have support as educational tools that might be useful in assessment as well.

In virtual reality scenarios, for example students are asked to make decisions as if they were a fictional character, such as a consultant, a general practitioner or simply themselves. They then have to make decisions and are faced with different kinds of situations from a consultation with a patient to a situation in an operating theatre (Cook & Triola, 2009). These are very time consuming scenarios to design and develop. To the present date there is no evidence that they can contribute towards the development or assessment of clinical reasoning, but this is still a recent innovation and more research is clearly needed to draw any conclusions regarding its impact on clinical reasoning or the features of effective VLP for clinical reasoning assessment (Cook & Triola, 2009).

B. Clinical Reasoning And The Curricula

The relationship between clinical reasoning and the curricula has been of interest to medical education researchers for many decades (Norman, 1989; Schmidt *et al.,* 1987; Groves, 2002; Koh *et al.,* 2008; Kieser *et al.,* 2009; van Gessel *et al.,* 2003).

The studies investigating the impact of the curriculum on clinical reasoning at an undergraduate level have mainly focused on the planned curriculum at the schools or programmes. Among these a common classification seems to be adopted, usually distinction between *Problem-based Learning* (Neufeld & Barrows, 1974) or *traditional* (Engel *et al*., 1991). This dichotomy is a common theme across many medical education literature reflecting a debate on-going for the last four decades on the benefits of PBL (Duffy, 2011; Cooke *et al*., 2007). Details about the curriculum being studied are often scarce, with only little information being provided about the rationale of authors behind the classification of a curriculum as 'traditional' or as PBL (Maudsley, 1999). This scarcity of descriptions makes studies interpretation very problematic, as under the classification many different realities can be found (McKimm, & Barrow, 2009).

Traditional curriculum seems to be a classification adopted to designate curricula that follows the Flexnerian model of pre-clinical –clinical sciences, with demarcated separation between the two, with only few, if any, integration between these two parts of the curriculum. The delivery of pre-clinical sciences is discipline-based with only minimal integration and makes use primarily of large group lectures and laboratory work

The PBL by opposition tends to be used to describe student centred curriculum whereby basic sciences would be learned by exploration and discussion of clinical case in small groups (Newman, 2004). These are very broad classifications, often used in a light manner; however, they point out key differences in the educational principles and practices of each programme or school allowing comparisons to be made as long as interpretation is careful and any excessive generalisation avoid.

One of the first reported studies on the impact of the curriculum on the development of clinical reasoning was conducted in 1981 (Neufeld *et al.,* 1981). These authors studied the development of clinical reasoning from a developmental perspective using a cross-sectional, randomly selected sample of medical students from three different years (Norman, 2005). A total of 22 students were also observed yearly

from entering clinical practice (at undergraduate level) until their first-year of postgraduate training. To test their clinical reasoning participants were asked to perform a clinical examination on a simulated patient with their thought processes measured by 'stimulated recall' of the videotaped encounter. This study concluded the clinical reasoning process remains relatively constant from undergraduate to postgraduate level, with only the measure of content in the students' hypothesis correlating with both outcome and educational level. This implies that medical students' knowledge is developing but the development of their clinical reasoning skills remains relatively constant during this period (Neufeld *et al.,* 1981).

Later, Patel and Norman compared the results of a traditional medical curriculum with the results of students in a PBL curriculum. These authors asked three groups students in each medical school (beginner- 1st six month of medical school; intermediate- in the middle of their training and final year) to provide diagnostic explanations of a clinical case before and after being exposed to basic science information. Overall students in the PBL curriculum revealed more "backward-directed" reasoning whereas students in the traditional curriculum revealed a "more forward-directed" (Patel *et al.,* 1991). PBL students also produced higher quality explanations using biomedical information, which can be a sign of a better ability to integrate new knowledge (Patel *et al.,* 1991). However, this study noted that along with these explanation there was a greater tendency to generate errors (Patel *et al.,* 1991).

Subsequently, one review concluded that there is no evidence that PBL improves generic problem-solving skills (Norman & Schmidt, 1992). However, the reason for this might be attributed to the fact that such skill is controversial and has been discredited in favour of domain and content specific problem solving skills. In a more recent review from the same author, while analysing diagnostic ability or clinical reasoning the authors found that in fact there was a small but significant effect of PBL (Norman & Schmidt, 2000 p. 721)**.** Additionally their study of PBL also concluded that over time PBL seems to increase knowledge retention and enhances both transfer of knowledge to new problems and the integration of basic science into clinical problems. Additionally PBL seems to increase motivation, interest and self-directed learning skills (Norman & Schmidt, 1992). These results were also supported by Hmelo-Silver's study. He compared 1st year medical students at two different medical schools, one with a PBL curriculum and the other with a traditional medical science base. This was a longitudinal study over a period of one year, where the

students were asked to provide pathophysiological explanations evaluated using protocol analysis. The results clearly favoured the PBL curriculum as fostering students ability to provide more integrated and complex explanations, and their ability to transfer reasoning strategies across learning environments (Hmelo-Silver, 1998).

Boshuizen initiated a study to validate a progress test designed to measure the growth of knowledge and clinical reasoning skills in a PBL curriculum. These results add to the body of evidence showing a high correlation between biomedical knowledge and clinical reasoning scores. Additionally they also pointed out that this correlation was much stronger than the correlation between clinical reasoning and the score in the behavioural sciences part of the test. This supports the idea that there no such thing as a generic problem-solving skill and that biomedical knowledge is crucial to clinical reasoning (Boshuizen *et al.,* 1997).

However Norman & Schmidt recommend that evidence for the influence of PBL should be examined carefully. Cognitive research should not be ignored as theory-based research is fundamental to the quality of any educational intervention. In curriculum interventions there are many different variables and any study that focuses on PBL as "a single intervention" and utilises the "usual cognitive and clinical outcomes" will arrive at the conclusion of minimal differences. PBL should be expected to foster outcomes that are not measured in traditional medical education curricula (Norman & Schmidt, 2000 p. 727).

More recently, Arts *et al.,* (2006) conducted an experimental study (with an experimental group vs. control group) to compare a "redefined PBL curriculum" (adaptation of PBL aimed at developing knowledge acquisition and diagnostic, problem solving accuracy) with a traditional curriculum in international marketing. They sampled 75 second-year marketing students. Students' problem-solving performance was measured using a multi-dimensional coding scheme, constructed on the basis of the literature on expertise. Their conclusions supported the idea that PBL fosters reasoning and problem solving. The PBL students outperformed the control students in various aspects of expertise when analysing the problem: *"The experimental group demonstrated a more extensive use of domain-specific concepts and inferences, more inductive reasoning, and both diagnoses and solutions of higher quality."* (Arts *et al.,* 2006 p. 84). However, their results also indicate that students in the PBL group use strategies closer to the non-analytical model of reasoning and/or forward reasoning (Arts *et al.,* 2006). These results are controversial, as previously

PBL has been associated with backward, deductive reasoning (Patel *et al.,* 1991). One possibility advanced by the authors is that this is due to the way in which the PBL process was planned, with increased focus on scaffolding, use of careful business case "templates" and guiding the steps to help students carefully consider options without jumping to conclusions, and these factors promoted a more expert-like type of reasoning (Arts *et al.,* 2006). Another possible explanation could be that this effect is due to the nature of knowledge and reasoning expected in the marketing field, as has been well established that is the particular nature of medical knowledge that confers to clinical reasoning its uniqueness (Patel *et al*., 1994).

Groves & Alexander in 2002 followed on work previously done by Bordage *et al*., (Bordage, 2007) using a set of 10 Clinical Reasoning Problems (CRP) with 290 medical students in the University of Queensland's MBBS Programme. CRP is an instrument based on key features approach where students are asked to indicate if new information supports or rejects their previous diagnosis hypotheses. The results indicated that this instrument is a valid and useful tool to assess clinical reasoning skills in individual students in PBL curricula. They gathered data on gender, age, nature of primary degree, selection criteria and academic performance in the first two years of the programme. Association of these variables with clinical reasoning skills was analysed using univariate and multivariate analysis. They concluded that gender (female) seems to be a good predictor of CRP scores independently of nature of first degree or academic performance (Groves *et al.,* 2002).

Support to the positive effect of PBL can also be found in a Hoffman *et al.,* (2006) study. These authors from the University of Missouri-Columbia School of Medicine (UMCSOM) used students' scores on the United States Medical Licensing Examination (USMLE) Step 1 and Step 2 to evaluate the effectiveness of the implementation of a problem-based learning (PBL) curriculum. Their results showed a significant positive effect of PBL curriculum for both Step 1 (more knowledge application focused) and Step 2 (more clinical problem solving focused). Their conclusion was that the PBL curriculum was preparing better students for residency and clinical practice (Hoffman *et al*., 2006).

Although not entirely consensual, Koh *et al.* conducted a systematic review of the effects of problem-based learning curricula, concluding that there is evidence of some positive effects after graduation in the cognitive domains. Coping with uncertainty and the use of evidence-based medicine had strong evidence in favour of PBL. However, both critical thinking/reasoning and problem-solving skills were only

weakly associated with the PBL curricula (Koh *et al.,* 2008). Nevertheless, similar results were found by Mamede *et al.* in a review of papers on innovation and PBL (Mamede *et al.,* 2006). These authors analysed six key papers on PBL research, its theoretical foundations and key findings and on-going questions in the field. They stated that research into PBL has been prolific and although not always consensual, results have emerged showing the advantages of the PBL curriculum are now much clearer than in the past. These authors call for more research to be conducted in the field to answer the still reaming questions, such as *"why and how do the theoretical constructs underlying PBL work in practice"* (Mamede *et al.,* 2006 p. 421). According to them this means that the research agenda should analyse particular aspects of PBL on defined variables rather than only focusing on the curriculum as a whole since such studies would contribute to a better understanding of the complex interactions taking place in PBL (Mamede *et al.,* 2006).

In 2011, Goss *et al.,* conducted a cross-sectional study aimed at comparing the impact of a PBL curriculum with a traditional curriculum on medical students diagnostic reasoning (Goss *et al*., 2011). They compared a total sample of 431, divided by students at the end of their first-year of clinical practice (4th year) and final-year medical students using the Diagnostic Thinking Inventory (DTI). This test is a self-reported measure of clinical reasoning characteristics, mainly flexibility[9] of thinking and knowledge structure. Their results suggest a differential impact of the curriculum on the development of diagnostic skills. Their results showed that there was an improvement over time on DTI scores in both curricula, which is consistent with the literature and the curricula expectations. They also found, that traditional curriculum students had higher DTI scores (both flexibility and memory structure sub scores) in both years, with differences in junior doctors being higher in the memory scores. The authors point to a few reasons for this unexpected result. First PBL students had one year of research away from the PBL structure, while the traditional curriculum students were continuously exposed to the same type of learning structure. Also the PBL students had less half-year exposure to clinical practice; however, as the authors point this would only impact the final-year medical students. For this group the authors argue that their results may be a consequence of PBL

9 "Flexibility in thinking refers to a trait that is beneficial in determining the right diagnosis during the diagnostic process." (Stieger et al., 2011 p. 2)

students having “ill-formed knowledge domains” due to the fact that during the PBL process they should learn clinical and biomedical sciences together (Goss *et al*., 2011).

As the DTI is a self-reported measure it is not possible to ascertain if these self-reports are translated into a better diagnostic performance from traditional curriculum students, or if this difference is being caused by a lack/over confidence of one group compared with the other. Prince *et al*. (2000) found that PBL students, although perceiving themselves as well prepared for practice, often are insecure about their knowledge base, this could cause the DTI scores to be lower, when actually their competency is as good as the traditional curriculum students. Although the authors do identify that this was overlooked in the discussion, therefore it is not possible to ascertain whether there were any indications of this within the sample studied.

In summary, although not always entirely consensual, the studies described seem to indicate that there is a differential positive effect of PBL on the development of clinical reasoning at an undergraduate level, nevertheless more than four decades after the first PBL course (in 1969 at McMaster University) clear evidence of such effect is yet (if) to be found. To understand possible reasons for this and what research has unveiled about the impact of this curriculum on the development of clinical reasoning, it is necessary to understand its historical development and research and how does it relate with the ‘expected’ impact on clinical reasoning. This will be the focus of the next section.

Additionally, now other curricular models that promote knowledge integration have been used by medical schools, mainly the integrated by systems model. This model contains some elements of the traditional curricula, sharing the focus basic sciences, but here these are presented in an vertically integrated structure with an emphasis on clinical relevance and not in a discipline based structure as in the traditional curricula (Dent & Harden, 2005). The effects of those have been largely ignored by research on problem-solving and clinical reasoning with only a few, like Schmidt *et al.,* (1996) considering this type of curriculum in comparison. Can this type of curriculum represent the best of both worlds for clinical reasoning development? How can clinical reasoning be assessed at an undergraduate level? What is the impact of the three types of medical undergraduate curriculum on student's Clinical Reasoning strategies at the beginning/end of their clinical phases? What is the impact of clinical practice?

The present research will make use of the unique opportunity to study these aspects in a traditional medical school with a predominantly passive teaching tradition (Coimbra, Portugal), a medical school with a different cultural environment and delivering a more contemporary systems based integrated curriculum (Nottingham, UK) and a school from a similar cultural environment which has adopted the problem-based learning strategy (Derby, UK).

C. Problem-Based Learning Curriculum

The next section will present an overview of the historical development and research into Problem-Based Learning and how its core principles relate with the previously described models for development of clinical reasoning. This will provide an important basis and justification for the research questions posed by the current work, mainly by those posed by study one: analysis of PBL sessions.

Problem-based learning can be used in literature to refer to a vast body of educational practices therefore it is important to differentiate between what Bereiter & Scardamalia (1999) define as lowercase 'pbl' and the uppercase 'PBL'. With the first being a vast range of educational methods focused around problems and small group discussions, that can take place independently of the adopted curricula of a school as a teaching and learning strategy. The other, PBL, referring to the curriculum model suggested by Barrows and Neufeld (Neufeld & Barrows 1974) based upon constructivist and cognitivist learning theories, where "the learner, the problem and the inquiry" are central to all the structures and procedures adopted by a particular school (Barrows 1986). Both in this and subsequent chapters PBL will refer to this uppercase definition.

C.1 Historical Development Of Problem Based-Learning Curriculum

Problem-based learning was pioneered by McMaster University Medical School in 1969 with a class of 20 students. This medical school had been set up only four years before, in 1965, as a response to the growing population with increasing healthcare needs and changes in the healthcare system. The recognised local clinical, educational and research expertise made this the perfect place for such a medical school (Neufeld & Barrows, 1974; Barrows & Tamblyn, 1980; Feltovich & Barrows, 1984; Hmelo-silver & Barrows, 2008).

This was a three-year "short" course (regular MD programmes in Canada are 4 years long), to which students were admitted on the basis of *"the quality of undergraduate applicants performance"* not on their area of study (Neville & Norman, 2007, p. 270) and with an increased number of mature students (Neville & Norman, 2007).

Recognising the increasingly short shelf-life of medical knowledge (Neville & Norman 2007) the "McMaster philosophy" focused on providing opportunities for student to

develop skills essential for the future careers as doctors, rather than just memorizing vast bodies of knowledge (Barrows, 1986).

This would be achieved by promoting a strong integration between basic and clinical sciences, by opposition to the more common medical curricula, at the time, where there was a clear separation between basic sciences and clinical experience (Jones et al. 2001). The ambitious, four main objectives to be achieved by the PBL curriculum therefore are: "*1. Structuring of knowledge and clinical context; 2. Clinical reasoning; 3. Self-directed skills; 4. Intrinsic motivation"* (Barrows, 1986 in Neville, 2009, p. 2).

PBL was so radically different from the most common medical curriculum that it had a "love it or hate it" effect among the medical education community. For example in many European countries undergraduate medical education would typically last from 5-7 years containing initial years of taught basic sciences followed by a clearly demarked period of clinical experience (Jones *et al*., 2001). In the following years after the McMaster programme, PBL generated a lot of discussion, with some adopting this model while others reaming sceptical of its results. Some of those who followed McMaster and implemented PBL in the 70s and 80s were, for example, Michigan State University (1972), Maastricht medical school (1974), Newcastle (Australia), Harvard University or Sherbrooke University (Barrows 1996).

The rapid expansion of PBL meant that different medical schools felt the need to adapt the model to their own needs and constraints. Expansion and hybridization also led to some misconceptions and misapplication of PBL arising, motivating several authors (Maudsley 1999; Woodward *et al*., 1997) to make efforts to classify and distinguish between PBL models (Savery 2006).

Barrows (1986) suggested a taxonomy of PBL curricula, where he clearly distinguished between PBL (or "close-looped Problem Based learning") and other curricula that uses cases or problems are used as "adjuncts to lectures" within a curriculum that is defined by the teachers but with some elements of group work and opportunities for self-directed learning (Barrows 1986). Savin-Baden (2000) proposed a differentiation of five models of PBL based different views on knowledge, learning and the role of the student adopted. Another model (Harden *et al.*, 1998) proposed 11 types of PBL based on the way in which programmes were focusing on generic concepts and rules or application examples where those rules and concepts were being applied. In this model there would be a continuum between a theatrical programme focused on generic concepts and rules and a 'pure' PBL programme where knowledge would be driven from examples of clinical situations or problems

model (Harden *et al.* 1998). Similarly Charlin *et al*., (1998) suggested another classification for PBL models using 10 dimensions. These authors used this framework to compare three different PBL medical schools, demonstrating that although different approaches were taken all programmes were still complying with the PBL philosophy by sharing a set of basic fundamental beliefs and practices (Charlin *et al.* 1998).

C.2 Core Principles of PBL

Barrows defined the basic fundamental beliefs/practices of PBL as: 1) learning should be student centred and self-directed; 2) occurs in small groups; 3) the teachers are facilitators of learning; 4) knowledge must be presented in a integrated manner from a wide range of disciplines and areas; 5) problems are both a way to organise focus and stimulate learning and the development of problem solving skills 6) assessment methods need to be aligned and reflect the ethos of the PBL curricula (Barrows, 1996). Although in its origins PBL was motivated more by a practical need and opportunity then by a theoretical basis, all these PBL elements seem to be aligned and supported by the views of constructivists such as Piaget, Dewey, Vygotsky, and others (Newman 2004). We will briefly outline what each of these encompasses.

Learning is student centred and Self-directed: The student assumes the central role in the curriculum and in planning their learning, taking responsibility and ownership for their learning (Barrows *et al*. 2000). In the origins of PBL students should not be given any learning objectives but the task to decide what they needed to learn. All the curriculum would be decided by the student based on their own perceived learning needs (Barrows 1996). This was possible in MacMaster because students were subject to a national examination at the end of their undergraduate medical education. This examination would assure the public and stakeholders of the competency of the future doctors. In countries were such examination does not exist such responsibility lies within the medical schools. This need to assure the public, stakeholder and regulatory bodies that the students had achieved the necessary competency to become doctors, made some PBL curricula to gradually evolved to a combination of students led learning objectives and curriculum led/mapped learning outcomes.

Small group session: the small group would create an environment for discussion to happen, while simultaneously requiring students to develop their communication and interpersonal skills (Barrows *et al*., 2000). Learning would take place with no lectures or didactic teaching taking during the four years of training. This environment promotes collaboration and peer-pressure between the students to achieve good results, to engage in self-study and to actively contribute to the group's discussions(Schmidt *et al.*, 2011, p. 795). Nowadays some of the PBL curricula include lectures, these maybe punctual sessions or actually timetabled along with the PBL group discussions (Wood, 2003).

Problems have a central role and learning based on problems would represent an alternative to previous disciplined-based artificial structures (Neville & Norman 2007).

Problems in PBL are at the core of the PBL curricula. These cases are anchors to motivate students to identify "what they need to know" (Barrows 1996) in order to solve the problem while simultaneously demonstrating the utility and application of the learned knowledge to the medical context (Savery & Duffy 1996; Wood 2003).

Importantly the PBL cases are artificially created scenarios, they should encompassed a few critical characteristics, that are not necessarily the ones from common cases in clinical practice but serve a specific pedagogic purpose within PBL discussions, this is often an aspects overlooked by researchers. The cases need to be ill structured to allow free inquiry and self-directed learning to happen, requiring intentionality and extensive preparation (Newman 2004; Savery & Duffy 1996). They also need to provide the basis for the identification of areas to be studied by the students, and simultaneously ensuring that everything that is learned during the self-directed learning can be applied back to the problem re-analysis and resolution. Also the cases must always require integration of different basic and clinical sciences (Newman 2004; Newman 2005; Barrows 1996), which in clinical practice may not be always the case, if one assume that pattern recognition is a strategy used by experts in common cases. These specific characteristics of PBL problems, allow the process case discussion to foster a deeper learning and increased situational interest (Rotgans unpublished data in Schmidt *et al.*, 2011, p. 799).

The **facilitator or PBL 'tutor'** is another critical element in PBL, the role of the facilitator is very different from a didactic teacher (Charlin *et al.*, 1998; Savin-Baden, 2003). In PBL facilitators do not provide answers and cannot be seen by the group as a source of knowledge. Their task is to guide the process of discovery. Facilitators or

PBL tutors should engage in what Barrows calls *"metacognitive communication",* it is their role to ask the students questions, the students *"should be asking themselves to better understand and manage the problem"* (Barrows, 1996, p. 5). The original McMaster curriculum took a radical view, defending that the facilitators should be experts in facilitation and not the subjects, so they would not be in a position to provide students with answers but rather guide their questioning and discovery (Barrows, 1996). This view has been challenged by some, ideally the facilitator would be an expert on both the facilitation process and the subject, these two characteristics would allow the tutor to have the necessary "cognitive and social congruence" to provide the students the flexible scaffolding that is associated with a positive effect on student learning (Schmidt *et al., 2011)*.

Finally, **assessment** and examinations should *"measure the progress towards the goals of PBL*" (Newman, 2004 p. 2). This is an essential aspect of PBL, and perhaps one of the most determinant aspects of its success or failure; if assessment is focused only on recall of factual knowledge, then PBL is likely to fail (Boshuizen et al. 1997). As students will focus on the knowledge they have to acquire they are likely to disengage from the PBL process and discussions, as those are much more time consuming then just simple textbook memorisation of facts (van den Bossche et al. 2000).

Assessment is still one of the major challenges faced by all schools adopting a PBL curriculum to find a valid, reliable and feasible combination of assessment methods that would be able to capture and assess the skills and attitudes that the PBL process develops (or aims to develop) in the students (Savin-Baden, 2004; Schwartz *et al.*, 1998; Engel *et al.*, 1991).

PBL is now approaching half a century and after many decades of debate this has become a more matured, more accepted and less radical, gathering extensive support from the community, but also different from the original model. In the year 2000, 60.9% of the Medical Schools in the US had adopted a PBL curriculum (Baker, 2000) making PBL one of the most common curriculum types in medical education (Wood, 2003). However, the initial concerns echoed about the competency of the "PBL doctors" led to an on-going debate and thriving research making PBL one of the most researched areas within medical education, answers are still to be found regarding the impact of these curricula models in specific competencies and skills, such as clinical reasoning(Norman & Schmidt, 2000). However, multiplicity of

practices within the PBL classification is highly problematic for research, especially because often details on the curriculum (e.g. assessments) are not reported, making discussion of different studies difficult, as in fact these maybe concerning different curricula under the same 'PBL umbrella'.

C.3 PBL Research

Problem-based learning has been intensively studied in the last 40 years (Neville, 2009). Since then this area has become probably one of the most researched areas in medical education, capturing both interest and critiques, with PBL being adopted by many medical schools all over the world. Even so, as in clinical reasoning research, results have not being unequivocal or consensual (Neville, 2009, Norman & Schmidt, 2000).

The pedagogical basis of PBL suggests that it should represent an improvement on traditional methods of medical education (Newman, 2005).

Learning knowledge in an integrated manner focused around clinical problems should facilitate the formation of this internal coherent integrated database of medical knowledge (Boshuizen *et al.,* 1995). Self-directed and student centred aspects of the PBL curricula would contribute to this aspect. Furthermore the use of problems would allow the students to start building their own internal databases of examples or possible patterns, which would facilitate the acquisition of new knowledge and its retrieval and mobilisation in clinical situations (Newman, 2004). However it is important to notice, as aforementioned, that there is a remarkable difference between a PBL case and a clinical case in practice (Kassirer *et al*., 2009; Kuipers *et al.,* 1988). PBL case/problems are artificial cases carefully designed to lead prompt the students to pursuit of certain knowledge and skills more then to achieve a diagnosis (Prince et al. 2005). Clinical practice the aim is to achieve a correct diagnosis and patient outcomes within the minimum amount of time 'navigating' the complex and multidimensional environment that is far more demanding than a classroom based discussion (Higgs *et al.,* 2008).

Nevertheless, initial assumptions were that the first (PBL) would promote the later (clinical effectiveness and clinical reasoning ability), but research has not been able to support these assumptions beyond doubt (Colliver 2000). Some authors have concluded that this type of curriculum shows an improvement in communication and interpersonal skills (Kaufman & Mann, 1998; Albanese & Mitchell, 1993; Schmidt *et*

al. 1987), whereas others have questioned its ability to promote knowledge acquisition and suggested it could lead to a deficit in knowledge (Hoffman et al. 2006; Watmough *et al*., 2006a,2006b).

More recently a study by Koh conducted a large meta-analysis comparing several different previous studies and concluded that there was a positive effect of this curriculum on social and cognitive attributes (Koh *et al*., 2008). However, other previous meta-analyses have demonstrated consistently small differences in outcome effectiveness for PBL compared to traditional medical education (Dochy *et al.,* 2003).

Additionally doubt has been cast on research methods and theoretical assumptions (Colliver & Markwell, 2007), with some attributing the non-conclusive results to "unreasonable excessive expectations" of the PBL process (Albanese, 2000). It has been alleged that comparisons have been made using instruments and methods that reflect the expectations and outcomes expected of the traditional curriculum, or that these methods are not sensitive enough to detect differences in interpersonal, problem solving and clinical reasoning skills. Hecker and Violato have recognised this problem and state that the results of licensing exams may not fully assess these important physician characteristics (Hecker & Violato, 2008). Also most of the studies have focused on outcomes such as national examinations, written assessments, practical skills assessments, but non, at least to the our knowledge, has used measures driven directly form clinical practice such as clinical outputs or (real) patient feedback.

Finally, the fact that the model of PBL has evolved into several different types of hybrid curricula, as mentioned before, implies that researchers must identify clearly the object of their study. More research into PBL is still a fundamentally important area of study that can reveal some of the underlying cognitive and affective effects of the method (Wood, 2006). The same can be said about both traditional and integrated curricula which are not single educational environments guided by a common set of principles and characteristics, since in practice they can differ from each other becoming more or less hybrid (Dent, 2005). Problem-based learning or pbl (lowercase) sessions can occur within traditional curricula, integrated curricula can have more or less expositive/traditional modules, to give two examples of hybridization (Harden, 2008). Therefore studying the impact of curricula on the

development of clinical reasoning demands a careful objective description of the curricula being compared (Muijtjens *et al*. 2007).

To the present date more than four decades after the first cohort of students graduated from McMaster University the question regarding the benefits of PBL is still not fully answered. Perhaps rather than focusing on overall benefits of PBL, researchers need to focus on identifying how particular aspects of medical competency are fostered by the different curricula and if the different pedagogical approaches have a differential impact on those. In the current research, we adopt this view by investigating the impact of different curricula in the development of clinical reasoning, a crucial element of medical competency.

Chapter 3: Context Of The Study – Analysis Of Three Medical Curricula

"Every great advance in science has issued from a new audacity of imagination."

(Dewey, 1929)

Chapter Summary

This chapter will begin by presenting a comparative analysis of the curricula of the three medical schools involved in the present research: Graduate Entry Medical school, University of Nottingham in Derby, the University of Nottingham Medical School in Nottingham, and the Faculty of Medicine in Coimbra, Portugal.

This analysis will be guided by the principles of Comparative Education and will make use of two recognised curriculum frameworks: SPICES (Harden, 1984) and the "integration ladder" models (Harden, 2000) to structure the comparison and present its results.

The first part of this chapter will present a brief description of key concepts and guiding principles of the analysis and describe in detail the frameworks used to compare the curricula. A second section will describe the curricula at the different schools. A final section will present the juxtaposition of the three curricula and their comparison.

The hidden curriculum[10](Muijtjens *et al*., 2007) and tacit assumptions regarding clinical reasoning development will not be included here, firstly because these would require a fine-grained analysis of teaching and learning practices, for example via observations of several educational moments within every school. This level of detailed analysis is out of the scope of the present research. Secondly, the hidden curriculum and its implicit messages are complex phenomena resulting from the interaction of many variables, often highly related to the culture of the educational environments (Muijtjens *et al*., 2007). However, if the formal curriculum is clear and properly aligned with the assessment strategies, its impact on knowledge acquisition and cognitive skills will hopefully also be aligned, or at least not a barrier, to the formal curriculum (Margolis, 2001).

---

[10] Hidden curriculum: describes the beliefs, values, aims, outcomes and practices that are not clearly stated as being part of the curriculum but are conveyed in it. It is also known as unintentional learning. (Margolis, E. ed., 2001. *The Hidden Curriculum in Higher Education.*, New York, USA: Routledge, Taylor & Francis Group.)

A. Comparative Education As Guidance To The Analysis

Comparative education[11] is an academic field of education dedicated to the study of education at comparing educational practices, processes and outcomes, at a country, system or institutional levels (Robertson, 2005). This field is often associated with comparison among geographical regions or international practices with studies such as PISA (OECD) among the most commonly associated with this field (Jarvis, 2010; Dale, 2005). However, there is much more to this field, with many other units of analysis being frequently used in comparisons other then regional/geographical locations, the curricula is one of those (Crossley, 2010; Broadfoot, 2010). Comparative Education studies are often complex large-scale projects involving multiple units of analysis, with curricula being one of those frequently considered. It is not the purpose of the present research to undertake in such an endeavour. Nevertheless, the model of comparison used in such studies provides a useful guide to identify the core aspects to be considered by those wanting to compare different educational realities or environments (Broadfoot, 2010; Bray,Admanson & Manson, 2007, Bray, 2005). The table below describes this model and how we made use it to guide our analysis. We will present the description of the different curricula after the description of the sources, and not as the first step in the comparison as commonly happens in comparison studies. As the criteria used to access information/sources has direct implication on the type and level of detail of description on the curricula presented.

[11] Disambiguation: Comparative education refers here to the academic field (with its own aims and methods), not to constant comparison method, a technique used within Grounded Theory as a method of analysing data in order to develop a grounded theory (for a more detailed view on Comparative Education please see references in text).

Table 3. 1: Bereday's Model of comparison in Comparative Education (extracted from (Bereday, 1964, p. 28)

Phases of comparison	Description	How it guided present analysis
Description	Presentation of key pedagogical concepts and data to be compared	Description and overview of the curricula will be presented based on the information from the sources.
Interpretation	Evaluation of the pedagogical data decision about weather or not can be included in comparison	Sources section will discuss the type of data to be included.
Juxtaposition	Establishing differences and similarities structured by a comparability framework	Analysis the three different curricula will be organised by the research questions, SPICES and Integration ladder frameworks will be presented.
Comparison	Simultaneous analysis and comparison of the juxtaposed pedagogical data in order extract conclusions	We will analyse the comparative table (phase III) and present the conclusions regarding the curricula highlighting its implications for the present research.

B. Defining The Key Concepts And Frameworks:

B.1 The SPICES model

The SPICES model was presented by Harden and colleagues in 1984. This model identifies two opposite groups of six core themes (Figure 3 .1) to be considered when analysing curricula. By establishing the location of each of these curriculum elements between the opposite extremes it is possible to characterise objectively the type of curriculum being used (Harden, 1984).

The themes presented in the SPICES model include *student-centred* in opposition to *teacher centred*. A student-centred curriculum is focused around the students' needs by opposition to teacher centred curriculum where clearly the focus is on the information the teacher finds important for the students to learn (Harden, 1984).

A second pair of opposite themes is *Problem-based versus Information-gathering.* A problem-based curriculum is a type of curriculum organised by clinical problems that incorporate elements from different disciplines and areas. Students learn by actively engaging in problem-solving activities that create structured opportunities for students to acquire skills and knowledge. By opposition an information-gathering curriculum is focused on providing the students with information they are expected to learn to successfully complete their degree (Harden, 2000).

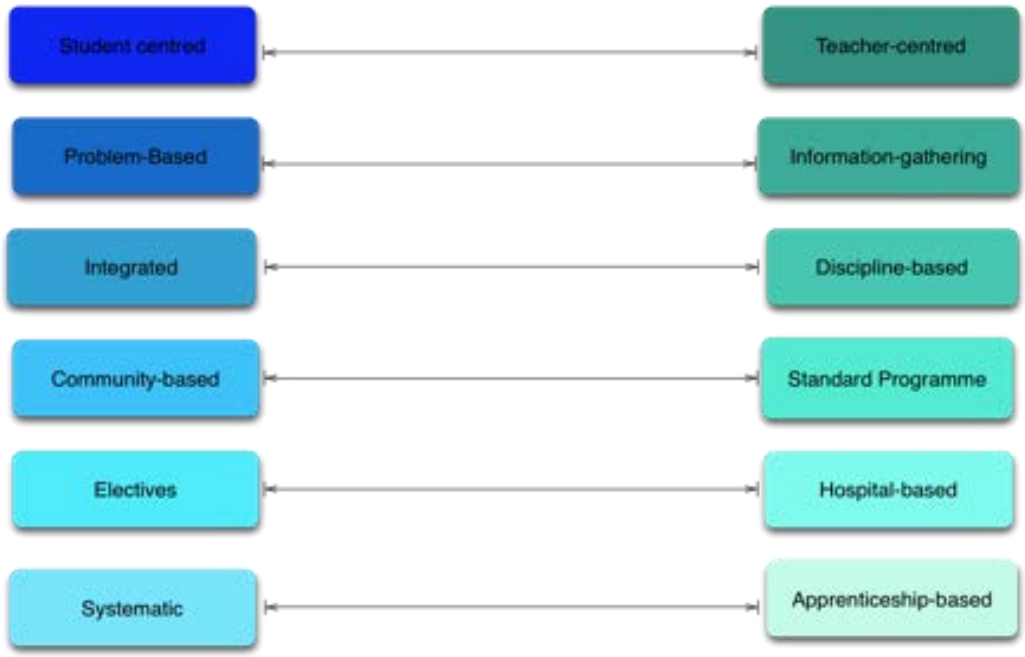
A third pair relates directly to the structure of the curriculum, being Integrated versus Discipline-based. A curriculum that is integrated is one where the core units of organisation do not reflect a single scientific discipline, e.g. biology, rather these core units are defined as units with meaning for clinical practice, e.g. body systems, diseases, or clinical presentations. Inside each of these core units knowledge of the different relevant scientific disciplines will be addressed in a collaborative way focusing on their contribution towards the particular unit e.g. the biology of respiratory diseases. Integration also implies that basic medical science teaching is clinically contextualised and that clinical teaching is supported by evidence from basic medical science.

Another two opposing themes are *Community-based* and *Hospital-based*. Medical schools can choose to be at either of the two ends or somehow in the continuum between the two. Usually this choice is guided by external factors, such as the needs of the community/countries/healthcare systems. This is not limited to the places where students experience clinical practice it has implications across the entire

curriculum, in the choices of modules structure and content (Harden, 2000). In a community-based curriculum social aspects of medicine, epidemiology, public health, prevention of deviant behaviours, community intervention among others, would stand as more relevant. On the other hand in a "pure" hospital-based curriculum specialities training would assume a more relevant role (Harden, 2000).

A final dichotomy is presented by the *Electives vs. Standard Programme theme.* Standard programmes are the ones where students have no or very little opportunity of being proactive in their learning process, as there are no opportunities for students to choose any of the learning opportunities they will be exposed to, all is defined previously by the school for all the students. An elective curriculum will allow students to have an element of choice via projects, special study modules, student selected components or elective experiences.

Figure 3 .1: SPICES model representation adapted from (Harden 1984).



B1.1. The integration ladder (Harden, 2000).

The integration ladder is another tool for curriculum planning and evaluation, focusing on the integrative aspects of the curriculum. The integration ladder model presents eleven levels, from a curriculum where disciplines are isolated (isolation)

from each other to a full integrated trans-disciplinary curriculum where content is organised not in scientific areas but around other meaningful units, e.g. Clinical problems.

Isolation is the first level and is characterised by a complete separation between subjects that are led by experts in their own field. Each unit of the curriculum is seen as a separate entity. Relationships between subjects are not covered in an explicit way. On the way to integration the next level of the ladder is ***awareness***; the difference between this phase and isolation is that some teachers may be aware of what others are teaching in their classes, some classes might be shared between two teachers of different subjects occasionally. The next level is ***harmonisation***, which describes a curriculum were teachers are responsible for different courses and communicate with each other, either informally or formally in committees. Consultation processes are in place, and although disciplines are still separate units there is connection between them. ***Nesting*** is the following level where "individual subjects or disciplines recognise the broader curriculum outcomes and relate their teaching programme to these" (Harden, 2000 p. 553), Borders between disciplines become more flexible. ***Temporal coordination*** follows, where although subject remains separate timetabling allows for different disciplines addressing similar topics to be taught on the same days close to each other. ***Sharing*** describes a curriculum in which some disciplines jointly plan the teaching programme, highlighting the relationships between concepts and skills in each of the disciplines.

Correlation is the stage where specific sessions or course units are introduced with the aim of exploring the relationships between other course units, e.g. an integrative module to explore the application of pharmacology to the treatment of respiratory diseases.

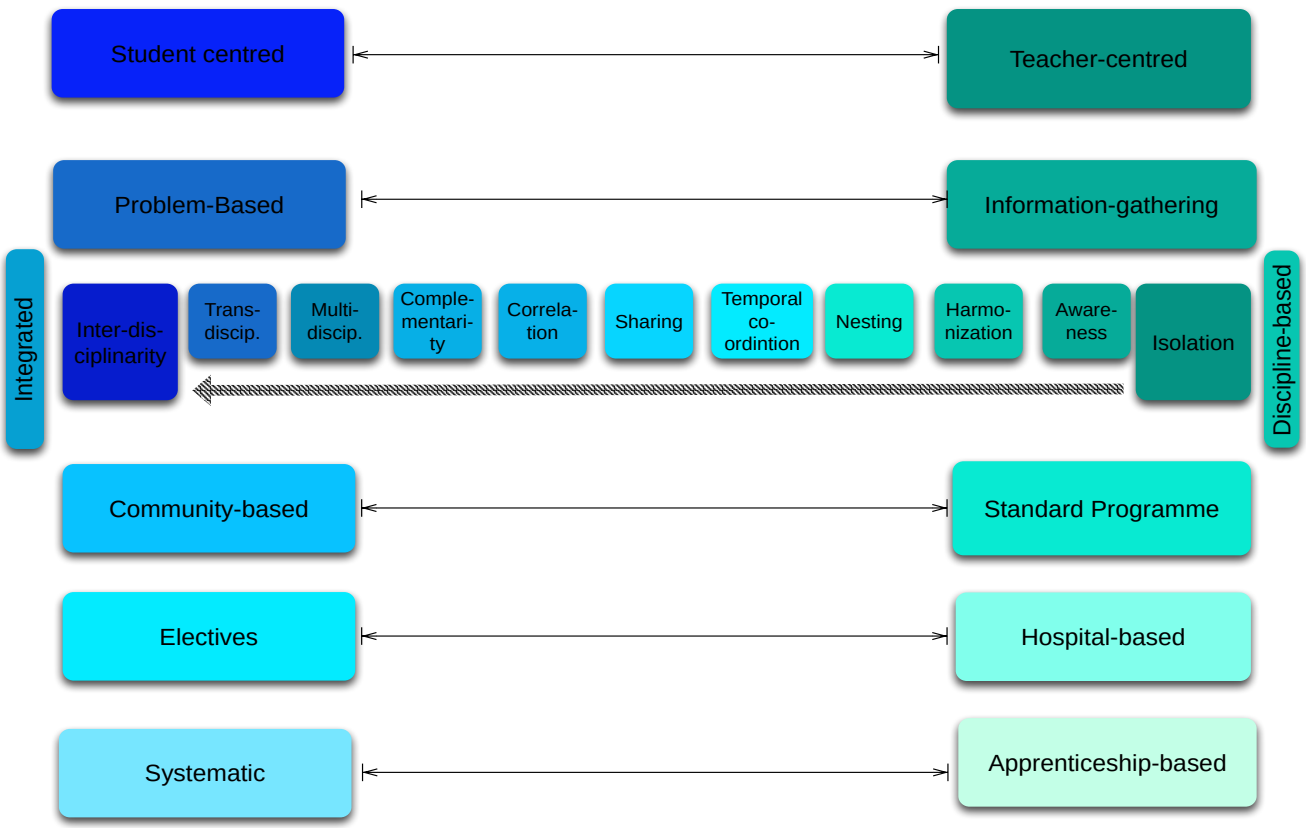
The Complementary curriculum stage consists of integrated sessions represent a major feature in the curriculum being as important as stand-alone disciplines. In a ***multi-disciplinary*** curriculum small number of single areas are joined together by a single course theme, problem or topic that focuses on students' learning needs with relevance for clinical practice. The ***Interdisciplinary*** stage is when the content of all or a majority of the single disciplines is re-combined into different units, without any mentioning of the names of the stand-alone disciplines.

Finally ***trans-disciplinarity*** consists of a curriculum with no separation among disciplines or their contents, and all is organised around the learning needs of the

student. Harden gives as an example, a curriculum organised into 113 clinical problems and task used as framework to guide students thought their final-year in Dundee medical school (Harden, 2000).

In summary this model provides a useful and detailed tool to characterise and compare curricula integration, as shown by It is also easy to integrate this model with the previously present SPICES model. This combined model will be the basis of the comparison presented in the following Figure 3.2.

Figure 3.2: Integration ladder visual representation integration between SPICES and Integration Ladder used in this study



B.2. Research Questions

What is the role of clinical reasoning in the documentation from regulatory bodies in the two countries?

This is a transcultural study comparing schools from two different countries, for this reason it is fundamental to ascertain if similar importance is given to clinical reasoning in the documents that regulate medical school practices, mainly Tomorrow's Doctors (GMC, 2009) and the Licenciado Medico em Portugal (Victorino, *et al.,*2005)

What is the importance given to clinical reasoning in the mission statement or values statements?

Value statements are expressions of schools values and principles that guide both the school outcomes and its culture. Identifying if and how clinical reasoning is mentioned will provide a measure of the importance the development of this skill assumes in each particular school.

What are the main differences and similarities between the three types of curriculum?

a) In their stated aims/teaching strategies/assessment structure?

To answer this question the SPICES and integration ladder models will be used (Harden, 1984, 2000) as a framework to juxtapose the curriculum models used in each of the three medical schools.

b) Opportunities for clinical reasoning development?

Curricula will be analysed to identify stated opportunities for clinical reasoning development in each school. Curricula will be searched for learning opportunities aimed at the development of clinical reasoning.

c) Clinical exposure and clinical practice opportunities?

When are these opportunities available for students and what are the aims and outcomes in each school. Aspects such as position in the curriculum, time and frequency will be analysed.

B.3. Sources

A few steps were taken in order to guarantee the accuracy of the comparison. All sources of information were reliable and equal in all schools, in that they were expected to contain information of a similar nature.

Sources of information about the schools were limited to primary sources, that are, documents produced by the schools themselves. This choice was made to ensure the coherence between the two countries in analysis. For the UK medical schools it possible to find documents from school evaluation, mainly the ones produced by the regulatory body the General Medical Council, with great detail regarding the teaching, learning and assessment practices in the school. That is not so for the Portuguese Medical Schools where no regular evaluation of medical schools is in place as a systematic procedure. Here the available documents would be reports form the Universities, based on either surveys of students or internal reports from the schools. Such disparity in the level of information does not contribute towards an objective comparison (Muijtjens, 2007). The primary sources, such as value statements, course handbooks, curriculum maps and other relevant information on school policy, were limited to publicly available documents, and documents available for students on the schools intranet.

Information made public by medical schools is intended to be representative of their educational practices, as is directed to stakeholders, regulatory bodies and prospective students. Schools' accountability is largely dependent on these documents, as it represents statements of their practices that can be verified by any of stakeholders, guaranteeing the accuracy of this type of information. Additionally future employers of their graduates and other institutions rely on these documents. For example exchange programmes with other universities in different countries, largely depend on publicly available curriculum descriptions and stated outcomes.

Information for students on the intranet, is one of the best sources of information about the curriculum, as these are the documents used by schools to make their students know "what they need to know". These provide a useful and more detailed view on the publicly available documents. Both groups of documents were collected via their websites and intranet.

C. Comparative Analysis: Juxtaposition

C.1. Regulatory Bodies guidelines

What is the role of clinical reasoning in the documents of regulatory bodies in the two countries?

In Portugal, medical schools, as universities, are regulated directly by the Ministry of Education, via the Higher Education Secretary. Additionally to open a medical school, approval is required from the medical association, *Ordem dos Medicos*. This institution's role is similar to the General Medical Council's in the UK, however with an important caveat, there is no regular evaluation of medical schools' practices. *Ordem dos Medicos*, is only responsible for authorising universities to award medical degrees by analysing submitted documentation and for the registration and supervision of all doctors who are practicing in Portugal; no regular evaluation of the quality of teaching and learning is made by this institution.

In 2005, a committee of all the medical schools in Portugal and the foundation of Portuguese Universities gathered to define a common set of outcomes to be fulfilled by all the medical graduates in this country. The documents produced *O Licenciado Medico em Portugal: Core Graduate Learning Outcomes Project* (Victorino *et al.,* 2005) followed closely the GMC's *Tomorrow's Doctors* (GMC, 2009), both in structure and type of contents included. The expressions related with clinical reasoning such *as clinical reasoning*, *critical thinking and reasoning* and *medical reasoning* are mentioned four times in this document. A first example is found on the section background and philosophy as quoted below:

"Teaching, learning and assessment methods should be student-centred, should encourage active learning and should promote **critical thinking and reasoning**." (Victorino et al., 2005 p. 64)

Another example can be found in the core document bringing clinical reasoning to the centre of the clinical competency and medical practice.

"Demonstrate proficiency in **clinical reasoning** through their ability to: recognise, define and prioritise problems; analyse, interpret, objectively evaluate and prioritise information, recognising its limitations; recognise the limitations of knowledge in medicine and the importance of professional judgment." (Victorino et al., 2005 p. 78)

Finally some direct consideration about the curriculum that should be used to promote the above is made:

"The curriculum should promote active, student-centred learning, should include evidence-based medicine and should include opportunities to demonstrate and practice **critical reasoning skills**." (Victorino et al., 2005 p. 81)

These statements place clinical reasoning at the centre of the outcomes expected from Portuguese medical graduates, where not only the importance of clinical reasoning is recognised, but also where a clear call to medical schools to use and promote learning opportunities that foster the development of this mental process is emphasised. In the equivalent document published by the General Medical Council (GMC) in the UK the expression "clinical reasoning" is not directly mentioned, however several very closely related concepts are highlighted. The foreword that introduces the documents starts by stating that:

"Doctors must be capable of regularly taking responsibility for difficult decisions in situations of clinical complexity and uncertainty." (GMC, 2009 p. 5)

This statement clearly addresses competency in clinical reasoning as central to future doctors, although focusing on its outcomes. Section eight of 'Doctor as Scholar' clearly states that graduates must be able to:

"(c) Justify the selection of appropriate investigations for common clinical cases. (...) (d) Explain the fundamental principles underlying such investigative techniques." (GMC, 2009 p. 28).

Finally section 14 of Doctor as a Practitioner states that students graduated in the UK must be able to

"Diagnose and manage clinical presentations […] (B) Make an initial assessment of a patient's problems and a differential diagnosis. Understand the processes by which doctors make and test a differential diagnosis." (GMC, 2009 p. 20)

In summary in both countries clinical reasoning is understood as a critical outcome to be developed by the medical students at the undergraduate level. Regulatory documents clearly recommend that medical schools promote the students clinical reasoning process at this level, developing adequate teaching and learning opportunities. However, what is the importance given to clinical reasoning in the different schools curricula? What are the main differences and similarities between the three types of curriculum?

C.2. The Curricula

The Faculty of Medicine of the University of Coimbra holds a six year medical degree, the three first-years of the degree are mainly dedicated to basic sciences, the following two dedicated to clinical placements in the local hospital and a full year of clinical practice with weekly seminars in the final-year (Figure 3. 3).

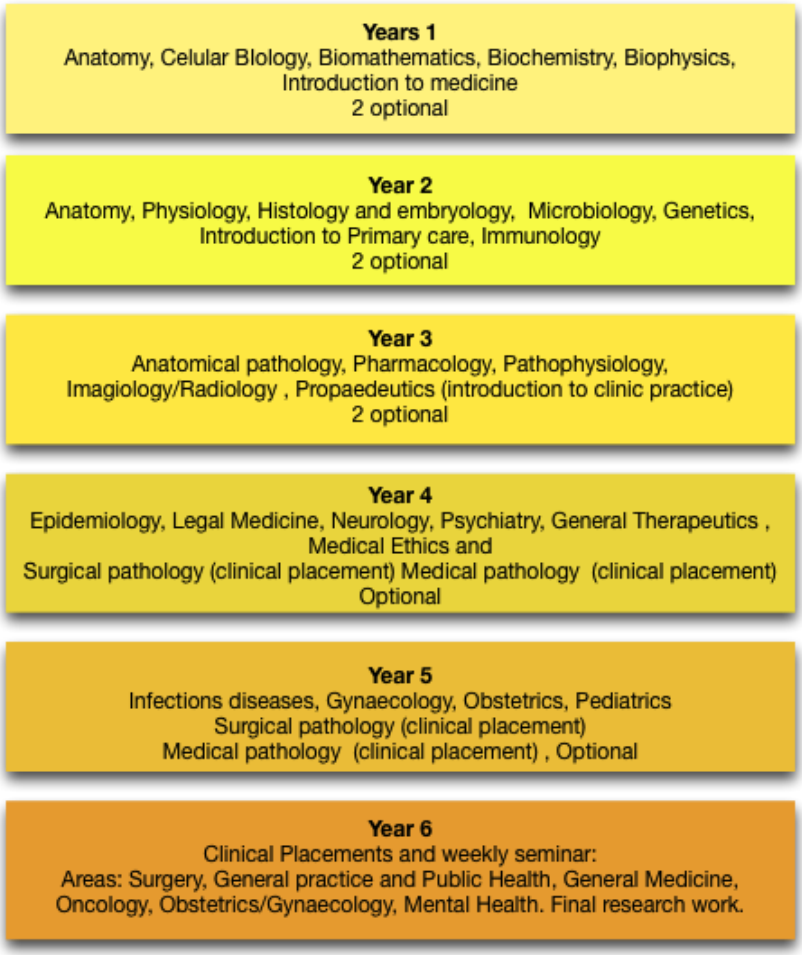
University of Coimbra Medical School was established in the beginning of the 13th century, in 1290. This school admits a total of 255 undergraduate students a year. Admissions are based on a national ranking of students based on their secondary education results (weight 50%) and the results of national exams in relevant subjects areas (e.g. Biology) (weight 50%). This is a highly competitive degree with the average score (in the last five years) of 18.7 points on a 0 to 20 points scale (national classification between the 95th and 99th percentile) (MCTES 2006).

The medical degree follows a traditional Flexnerian structure, combining three years of basic medical sciences with three years of clinical sciences & practice. During the first three years the students engage in lectures, laboratory work and some small group teaching sessions covering areas like anatomy, biochemistry, pharmacology, physiology, pathology. The following three years are dedicated to clinical rotations in the local university hospital combined with lectures and seminars regarding the different medical and surgical specialties[12]. In summary we are measuring and comparing clinical reasoning in a predominantly passive 'pre-clinical – clinical' teaching tradition (Coimbra, Portugal), a medical school delivering a science focused integrated systems based curriculum (Nottingham, UK) and a graduate entry medical school that adopts a problem-based learning strategy (Derby, UK).

In this study we have made use of the unique opportunity provided by access to these different medical curricula to gain an understanding of their influence on differences identified in the students clinical reasoning outcomes and strategies.

12. Information available at http://www.fmuc.pt [accessed 10 October 2011]

Figure 3. 3: University of Coimbra (FMUC) Medical Degree course structure adapted from documents available on http://www.uc.pt/fmuc

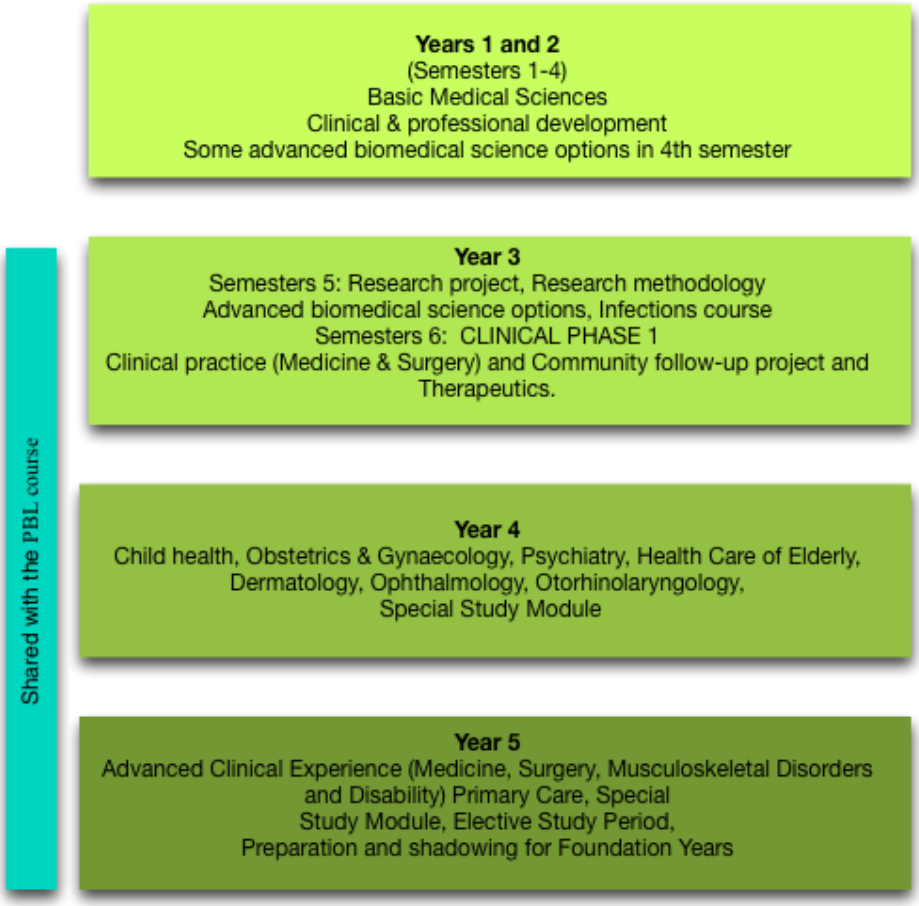


The University of Nottingham has two medical schools, one a five-year BMedSci-BMBS programme with an additional research focus and a BM BS Graduate Entry Medicine Course. The students in this programme have two and half years of systems-based integrated basic sciences teaching followed by a similar amount of time in clinical placements (Figure 3.4).

The other medical school at the University of Nottingham is a graduate entry school, that holds a four-year programme, with the two first-years being dominated by 4-5 hours per week of PBL sessions, few lectures, clinical skills training and some shadowing in community-based sites. The two following years are dedicated to clinical placements along with their colleagues from the five-year programme.

The **BMedSci/BMBS medical degree** is an integrated by systems based curriculum with a research project on the Honours' year (Figure 3.4). This degree includes a series of lectures, small group teaching, laboratories work and clinical skills training during the first four semesters, covering three main areas: basic biomedical sciences, clinical and professional development and advanced biomedical sciences. Semester five is dedicated to a research project along with research methodology, advanced biomedical courses and an infections course (University of Nottingham, 2011a). After the successful completion of these, the students join their colleagues from the GEM course and enter five semesters of clinical practice. The clinical phases include clinical rotations for several different clinical placements in hospital and the community health services in the region, covering a range of medical and surgical specialties.

Figure 3.4: BMedSci/BMBS course structure extracted from Course handbook (University of Nottingham, 2011a)



Graduate Entry Medicine Course (**BMBS)** is held at Graduate Entry Medical School in Derby. This school, opened in 2003, admits 90 students a year from many different backgrounds, with no restriction on heath related professions. The selection of the candidates is done via a GAMSAT[13] examination plus a structured interview process. This is a very competitive process and only approximately 7.5% of the candidates applying are admitted (GAMSATUK, 2011). One of the elements in this recruitment process is an evaluation of the willingness of the candidates to take part in a PBL curriculum.

Moreover, this school provides a four year medical degree, adopting an Problem-based learning curriculum (Figure 3.5). With the first 18 months being dominated by a PBL sessions followed by 5 semesters of clinical rotations throughout hospitals and GP practices around the in the region. The PBL sessions are organised around 11 main curricular themes[14]. Theme length may vary according to the themes specification, however all include a combination of lectures, clinical skills practice (attended in small groups on available slots), basic and clinical workshops (BCS WS, also attended in small groups), professional development sessions and clinical cases discussions (PBL sessions). Although the students have the responsibility of defining their own learning outcomes for the PBL sessions, there are overarching outcomes

---

13. "GAMSAT evaluates the nature and extent of abilities and skills gained through prior experience and learning, including the mastery and use of concepts in basic science, as well as the acquisition of more general skills in problem solving, critical thinking and writing. If your first degree is in a non-scientific field of study you can still sit GAMSAT and be chosen for admission to the graduate-entry programmes. A science background is not a prerequisite and academic excellence in the humanities and social sciences is encouraged and recognised. However, it must be stressed that success in GAMSAT is unlikely without knowledge and ability in the biological and physical sciences." Information available on http://www.gamsatuk.org/index.php?option=com_content&view=article&id=12&Itemid=48 [accessed 08 October 2011]

14. Structure, function and defence; Respiratory sciences, Respiratory sciences, Limbs and Back, Alimentary, Endocrine, Personal and professional development 1, Urogenital; Neuroscience, Integrative, Personal and professional development 2. (University of Nottingham 2011b)

for the curricular themes are defined by the school. Each week the students spend 4.5 hours discussing clinic cases selected according to the themes. These 4.5 hours are divided into 3 sessions, broadly consisting of the following:

First session - Case presentation, history and learning areas the students need to develop during the case.

Second session - Further exploration of the case, dealing with results of history, physical examination, clinical investigations.

Third session - Conclusions, management plans, treatments options and summary of learning areas.

For each case a facilitators guide is produced, that guide will help a group of clinical and non-clinical facilitators guide the students' learning through the course of these sessions. Facilitators' guides cover both the content, information about the case, and the process of the PBL sessions, having clear information about key learning areas (biomedical and clinical) that the students should achieve and possible questions to be asked in different moments of the case. These guides are crucial to the success of the PBL sessions. They guarantee consistency and provide direction across facilitators, groups and cohorts.

The curriculum adopted by this school is aligned with the pedagogical principles of PBL although adopting a hybrid format with the inclusion of lectures, clinical skills and basic/clinical sciences workshops, as to provide students with a diverse range of learning opportunities outside the PBL discussions.

Figure 3.5: GEM course structure available from (University of Nottingham, 2011)

School of Graduate Entry Medicine & Health
Year 1 timetable for A101, 2008-09

Start time	Monday	Tuesday	Wednesday	Thursday	Thursday	Friday
09.00	Lecture slots (4 max.)	PBL session 3 (1.5 hours)	Lecture slot	Lecture slot	GP Visit 7 visits inc. CFU Lecture slots (3 max.)	BCS WS 2 3rd session
09.30						
10.00				Lecture slot		
10.30		Slot for weekly Q&A session	BCS WS 1 1st session			
11.00				Lecture slot		PBL session 2 (2 hours)
11.30		PBL session 1 (1 hour)				
12.00						
12.30						
13.00			BCS WS 1 2nd session			
13.30		Clinical skills session		BCS WS 2 1st session	BCS WS 2 1st session	Clinical skills session
14.00						
14.30						
15.00			BCS WS 1 3rd session			
15.30		Clinical skills session		BCS WS 2 2nd session	BCS WS 2 2nd session	Clinical skills session
16.00						
16.30						

Example of weekly timetable

First 18th months
(Semesters 1-3)
PBL sessions
Clinical skills (ECE)
Personal and Professional Development (PPD)

Semesters 4
CLINICAL PHASE 1
Clinical practice (Medicine & Surgery) and Community follow-up project and Therapeutics.

Year 3
Child health, Obstetrics & Gynaecology, Psychiatry, Health Care of Elderly, Dermatology, Ophthalmology, Otorhinolaryngology,
Special Study Module

Year 4
Advanced Clinical Experience (Medicine, Surgery, Musculoskeletal Disorders and Disability) Primary Care, Special
Study Module, Elective Study Period,
Preparation and shadowing for Foundation Years

Shared with the BMedSci/BMBS course

C.3. Juxtaposition Table:

Table 3. 2: Comparative Table Of The Three Curricula Being Studied

Aspects Compared	FMUC	BMedSci/BMBS	GEM
1.Student-centred vs. Teacher-centred	Key aspect considered in this classification was the learning outcome definition, mainly if those are defined from a student or a session/teacher perspective (Harden 1984). Selected examples are included below.		
	Teacher-centred: Learning objectives are defined based on the knowledge to be delivered or materials to be taught during the session. Commonly found expressions are "expose students to knowledge of…' "add to students knowledge of…". Also common in the documents analysed are list of knowledge topics to be covered during the different modules [1,2.]	Some elements of student-centred and some elements of teacher-centred: At the curriculum map the learning outcomes are defined from the students perspective "What the doctor should know, do" or " [students should be able to] deal with uncertainty" "exhibit creativity/resourcefulness" [6 p. 22.] Although there is no scope for students to define their own learning objectives, those are set by the teachers/module leads/programme managers [8].	Mainly student-centred learning : learning outcomes are defined by the students in the PBL discussion groups, with a support form the facilitator [13] with curriculum map being defined from the students perspective. "What the doctor should know, do" [7 p. 22] Some didactic elements are included, such as clinical skills and workshops but seem to be defined from the students' learning needs perspective.
2. Problem-based vs. Information-gathering	A problem-based curriculum is a type of curriculum organised by clinical problems that incorporate elements from different disciplines and areas. An information-gathering curriculum is focused on proving the students with information they are expected to learn to successfully complete their degree (Harden, 2000).		
	Information gathering: By analysing the programme timetable, learning objectives and assessment practises, it is possible to see that each module corresponds to a unit of knowledge (e.g. Anatomy, Physiology)	Mix between problem-based and information gathering: The programme is organised by modules reflecting body systems, with variability between modules. Although it is possible to find module that adopt small group	Problem-based: Initial years of the curriculum are based around PBL discussions with some lectures and clinical skills training [15]

	with knowledge based outcomes and modular assessments. Also in the value statement of the programme it is possible to "see" the centrality of knowledge. *"By the end of the course students should demonstrate knowledge of basic and clinical sciences as well as skills needed to practice Medicine under supervision"* [3]	and lowercase pbl (as mentioned by Bereiter and Scardamalia, 2000 in M. Newman, 2003, p. 1) as the teaching/learning method [8].	
3. Integrated vs. Discipline-based	Isolation: Disciplines are a transposition of scientific fields, experts are the leading professors and each of he disciplines is independent of all the rest of the curriculum [1-4].	Between Multi-disciplinary and Inter-disciplinary: Years 1 and 2 In the clinical years, the course structure is organised in generic themes that group information from several different disciplines such as Human Development Structure and Function; Cardiovascular, Respiratory and Haematology or Alimentary System & Nutrition. However there are still some single subject modules such Public Health and Epidemiology or Behavioural sciences although representing a majority in the curriculum. Clinical placements are divided by specialities, although exploration of relationships and connections between different placements are encouraged [6].	Trans-disciplinary: The course is structured around body systems, each cycle corresponding to a group of weeks and a body system. Each week in a cycle is centred around a particular clinical case. Students discuss that case of the week in the PBL sessions (integrating knowledge from across a range of biomedical sciences), have clinical skills training opportunities relevant for that case and when necessary lectures in fundamental aspects related with the case and the cycle/system. In the clinical years, placements are divided by specialities, although hospital exploration of relationships and connections between different placements are encouraged [7,13].
4.Community-based vs. Hospital-based	Classification below was driven by an analysis of the courses timetables, and programme handbooks.		
	Hospital based, with some placements in general practices [2,3.]	Combination of hospital and community placements stated in the course handbook [9.]	In the first two years the students have the opportunity to shadow general practitioner. Last two years placements are a combination of

			hospital and community placements. This information is stated in the course handbook and the schools website [6, 12.]
5. Electives vs. Standard Programme	Standard programme with two optional courses to be chosen from predefined list year. All of the options offered are taught modules covering a range of non-core areas such as applied molecular biology, Health economics and management in healthcare, Neurosciences or foreigner languages (Medical English). Additionally there are several research projects to be developed across the 6 years, where students have the opportunity to choose themes [1].	Number of electives and weight in the curriculum increases over the years. Years one and two advanced biosciences options on last semester; year 3 1st semester is entirely dedicated to a research project chosen by the students; years 4 and 5 students have special study modules from a list of 44 different possible clinical placements (e.g. Genetic Cancer & Colorectal Surgery, Emergency medicine) in different place across the region. These placements last for 9,5 weeks in year the two final-years [7,11].	First 18th semesters are relatively standard last two include years 9,5 weeks of special study modules, shared with the BMedSci curriculum students (see box on opportunities the left). In the first 18th months of PBL students are given opportunities to choose some of they're learning opportunities, however the core curriculum is standard [6, 7,12].
6. Systematic vs. Apprenticeship-based	Scientific area based with final-years clinical placements are divided by medical specialities [2].	Mixed - Systematic with a few being modules corresponding to scientific areas. Systematic modules examples "Cardiovascular, Respiratory and Haematology; Renal and Endocrine Systems "[11a] Scientific modules examples: "Public Health and Epidemiology; General and Biochemical Pharmacology" [11a]	Systematic (cycles are organised in function of the body systems). Examples "GEM Alimentary System; GEM Cardiovascular sciences; GEM Endocrine System" [12a]
Other key aspects for the present study:			
Assessments	Isolation: Each discipline lead is responsible for designing and marking their disciplines assessments. The students sit an exam per discipline, their marks are given individually	Between Multi-disciplinary and Inter-disciplinary: Reflects the teaching structure there is a clear policy regarding assessment in place. Assessments are the responsibility of the	Trans-disciplinarity: Students are assessed thought a portfolio analysis and a summative assessment covering all the basic and clinical sciences at the end of year one and at the end

	and questions cover only the content of that discipline. There is no specific policy on formative assessment, being the responsibility of the discipline to decide the type of assessment. Additionally clinical placements are assessed by the clinical teachers according to their choice of assessments methods [1,2]. These usually have a written examination part with MCQ and a mini-OSLER's, history case discussions, performing of specific clinical skills. Some have introduced MIni-CEX's or OSCE stations, however no clear policy is in place to guarantee coherence between assessments.	module conveners, they might ask other lecturers that took part in the module to write some part of the questions as long as coherent with the assessment policy. Additionally to the modules assessments, students have OSCE's, complete portfolios and logbooks reporting their clinical experiences. Assessment covers all knowledge, skills and attitudes and behaviours. Information is outlined in the scheme of assessment and progression [7].	of the 18th months of PBL curriculum. The assessments are not divided in areas or systems but encompass questions that aimed at application and integration of knowledge into problems and clinical relevant situations. Clinical skills are assessed continuously throughout the 18-month course. Formative assessment is frequent and guided by the same norms that final assessments are. This school as introduced a clinical reasoning assessment as formative for the first-years and summative for the final-years [6].
Clinical reasoning opportunities	This classification was based on an analysis reading of courses handbooks, curriculum maps, and schools websites.		
	Clinical reasoning is not directly mentioned as an outcome in any of the disciplines; 4th, 5th and final-years clinical placements are the only component in the curriculum where an intention to develop clinical reasoning can be identified although is not directly stated as such. [2,3]	Reference to clinical reasoning cannot be found in the aims/outcomes of modules, however some references to it can be found in course materials for students, for example in the analysis of cases (e.g.A1HEC case reports analysis). There are e= indirect references to it, mainly refereeing to differential diagnosis in the outcomes of clinical placements (last 5 semesters) [7].	Some PBL cases explicitly focus on clinical reasoning development and /or differential diagnosis. Reference to clinical reasoning development can also be found, mainly referring to differential diagnosis in the outcomes of clinical placements (last 5 semesters) [6]
Clinical exposure	Classification based on information from the courses timetable available from courses handbooks and schools websites.		
	3 years of clinical placements, with a few	Two years and half (5 semesters) of clinical	Two years and half (5 semesters) of clinical

	opportunities of shadowing and observation of clinical procedures in second and third years [1-4]	placements and some opportunities for shadowing during the first three years [7, 8-12]	placements and opportunities for primary care placements during the first 18th months [6, 8-12]
Others contextual aspects:			
Admissions	Classification based on information on application and entry requirements available from national or the universities websites.		
	Very competitive National competition on the basis of national exams and secondary exams only: Students are allocated to the university based on their scores. Scores are calculated based on 50% results on national exam on especial disciplines (biology+chemistry/maths/physics) and 50% others secondary marks. The minimum score to be admitted in to FMUC in 2009 was 18,2 (highest score was 19.8) on 20 points scale [4].	Very competitive *"A in chemistry and biology at A level; third A level at grade A in any subject except general studies and critical thinking; at least six GCSEs at grade A including chemistry, physics and biology or double science; GCSE grade B in English and maths. Graduates: 2:1 degree in a science-related subject; A in chemistry and biology at A level; third A level at grade A in any subject except general studies and critical thinking. All candidates must sit the UKCAT test"* [9]	Very competitive *"Candidates who will have the minimum of a lower second-class degree must apply through UCAS in the usual way to course code A101 (the closing date is 15 October 2011). In addition you must have sat the GAMSAT examination (which is designed to ensure that entrants have the requisite knowledge and reasoning skills)"* [10]
Sources:	Course handbooks: 1)http://www.uc.pt/fmuc/ensino/mim/contentorDocumentos/MapaOp 2) https://intra.fmed.uc.pt/ Websites: 3) http://www.uc.pt/fmuc 4) http://www.sg.min-edu.pt/pt/acesso-ensino-superior/# Others: 5) Victorino, R.M., Jollie, C. & McKimm, J. eds., 2005. O Licenciado Medico em Portugal: Core Graduate Learning Outcomes Project, Lisboa, Portugal: Faculdade de Medicina de Lisboa.		Course handbooks: 6) University of Nottingham, 2011b. Graduate Entry Medicine: Course Handbook, Nottingham, UK: University of Nottingham. 7) University of Nottingham, 2011a. Medical School Course Handbook: A100, A101, A300, A105 A104. Medicine, Nottingham, UK: University of Nottingham. Websites: 8) http://www.nle.nottingham.ac.uk/year1/index.php?session=2011/12 9) http://www.nottingham.ac.uk/ugstudy/courses/medicine/bmbs-medicine.aspx 10)http://www.nottingham.ac.uk/gem/prospective/graduateentrymedicine/courseoverview/clinical-training.aspx 11) http://www.nottingham.ac.uk/ugstudy/courses/medicine/bmbs-medicine.aspx 11a) http://www.nottingham.ac.uk/ugstudy/courses/medicine/medicine-with-a-foundation-year-bmbs.aspx 12)http://www.nottingham.ac.uk/ugstudy/courses/medicine/graduate-entry-medicine.aspx 12a) http://www.nottingham.ac.uk/ugstudy/courses/medicine/graduate-entry-

		medicine.aspx 13) http://www.nle.nottingham.ac.uk/year1/index.php?session=2011/12 [facilitators guides] Others: 14) GMC, 2009. Tomorrow 's Doctors, Available at: http://www.gmc-uk.org/education/undergraduate/undergraduate_policy/tomorrows_doctors.asp. 15) GEM Year 1 Timetable: http://www.nottingham.ac.uk/gem/images-multimedia/misc/timetabletemplateyear1.jpg

Notes: The table below presents the information and description of the aspects considered within the different criteria used to classify the curricula. Examples are used when possible, however in some instances that classification was driven by large or multiple pieces of information (e.g. programme timetable) which could not be feasibly added to the table.

D. Conclusions

In summary a few points can be drawn from the above comparative study. First by the previous analysis we can conclude that the three curricula are good representations of the most common types of curricula in Medical education, that being: traditional 'traditional' curriculum (FMUC), Systems based Integrated Curriculum (BMedSci) and a PBL curriculum (GEM), their characteristics map onto the typical characteristics of these types of curricula described in the literature (Dent & Harden, 2005; McKimm, 2003; McKimm & Barrow, 2009; Dennick, 2008).

Second, regulatory bodies in both countries have established clinical reasoning as a fundamental outcome for undergraduate medical education agreeing on in its centrality to clinical competency. However, in the Portuguese medical school this intention does not seem to be applied into curriculum design during the first three years of the programme. In this school clinical reasoning is a tacit outcome, it can be found under the large umbrella of clinical competency mentioned in some of the disciplines but there is no explicit reference to it in the outcomes of the general curriculum, apart from in some clinical placements.

The opposite case is seen in the Problem-based learning (GEM) school, where clinical reasoning assumes a much more central role not only by being an expected outcome from the PBL sessions but also being part of the school's final assessments.

In the BMedSci/BMBS course, an intermediate situation can be seen. On the one hand the references to clinical reasoning in the general outline of the curriculum and curriculum structure are scarce, being closer to the FMUC in this respect. On the other hand in some course materials available online for students, some references to it can be found. In all three curricula clinical placements are a) organised in a similar way and contain similar activities and areas b) the place in the curriculum where clinical reasoning references tend to be more frequent.

Our study will investigate the impact of these different educational practices on clinical reasoning development by comparing the performance of two cohorts of students on a Clinical Reasoning Test, one cohort with only some clinical exposure and a second a few weeks before graduation in each of the three schools. The results of this study will be presented in chapters 7.

Chapter 4: Analysis of PBL Sessions (Study 1) – Methodology

"When I use a word, Humpty Dumpty said in rather a scornful tone,
it means just what I choose it to mean — neither more nor less.
The question is, said Alice, whether you can make words mean so many different things.
The question is, said Humpty Dumpty, which is to be master—that's all.
Alice was too much puzzled to say anything, so after a minute Humpty Dumpty began again.
They've a temper, some of them—particularly verbs, they're
the proudest—adjectives you can do anything with, but not verbs—
however, I can manage the whole lot of them! Impenetrability!
That's what I say!" (Carrol, 1871)

Chapter Summary

This chapter, as Lewis Carol quote reveals, is about words and their meaning, and mainly describes how in the present research we dived into the fascinating world of words to unveil their meanings and to use them to access some cognitive processes involved in clinical cases discussions.

The current chapter describes the methodology used in study 1 (Analysis of PBL Sessions) this is one of two studies that make part of the current research. This study is aimed at contributing to what is currently known about the cognitive processes used by medical students to approach, discuss and 'solve' clinical cases at an early stage of their medical education (first and second-years). This was conducted in a PBL setting, this context provides an extraordinary opportunity to observe and collect data on these processes, by requiring the students to share reasoning and acquired knowledge among the group, these sessions stimulate verbalisation of such cognitive processes.

A first section of the present chapter presents a description of a methodology from linguistics to the study of cognitive activities taking place in PBL discussions: corpus analysis, followed by a description of a methodology used to identify the content of the discussions: electronic content analysis. We will discuss the contributions of these approaches, and justify its need for the current research.

A second section will present a validation study, conducted in order to assert the applicability and usefulness of corpus analysis to unveil some of the cognitive activities taking place during the PBL discussions.

A third and final section will describe study 1 detailed methodology, outlining how corpus analysis was applied in the wider context of the current research. We will outline the philosophical approach underpinning this study, study purposes and research questions, sampling and ethical considerations, data collection and data analysis procedures and criteria to ensure rigour of the study.

This last section will provide the necessary basis to understanding of the next chapter (chapter 5) where the results of this analysis will be presented and discussed.

A. Electronic Text Analysis of PBL Sessions

A.1. The Need for a New Methodological Approach

The present section attempts to answer the questions: Why was the new approach described in this chapter needed? To what extent does it contribute towards the development of this research field? The answers to these questions justify the need, purpose and application of the methodology and are described in the following sections.

Problem Based Learning (PBL) has been the subject of a considerable amount of educational research, particularly with respect to evaluating its effectiveness in comparison to other, more traditional, forms of medical education.

Wood has argued that there is a methodological problem surrounding PBL studies. Performing outcomes-based research is beset with difficulties because of the large range of confounding variables scattered among the many curricula models (Wood, 2003) and because often there is a need for disambiguation when referring to curricula types. For example several authors have pointed out the 'PBL versus traditional' dichotomy may not be so distinct among the range of hybrid curriculum types now common in medical education (Newman, 2005; Harden and Davis, 1998).

More importantly, the assessment instruments that have been used to assess the knowledge, skills and attitudes of learners may be too crude to reveal the cognitive and professional advantages predicted by the 'theory' of PBL. These are often adaptations of assessments used in science based curricula, which do not allow PBL students to fully demonstrate the additional interpersonal, problem solving and clinical reasoning skills they might have acquired (Hecker and Violato, 2008).

Relatively little attention has been paid to the fact that students who engage in PBL talk to each other for 3 or 4 hours per week which can lead to hundreds of hours of interpersonal discussions over a complete PBL curriculum (Koschmann and MacWhinney, 2001). Theory would predict that this level of discourse should lead to changes in the use of technical vocabulary, interpersonal communication skills, the use of problem solving strategies and the development of clinical reasoning skills (Kamin *et al*., 2001, Yew and Schmidt, 2007), although little research has been carried out to investigate it.

In order to understand the impact of PBL discussions and the full extent to which learning occurs within these discussions research has to "*open the black box*" and

focus on the actions taking place during the sessions (Yew & Schmidt, 2007). As Loftus' work highlights, there is a close link between language and clinical decision making that ought to be explored (Loftus, 2006).

A number of studies have carried out observations of the PBL process. For example de Grave *et al*. (2001) and Schmidt *et al*. (1989) have demonstrated how small group discussions and collaborative learning processes taking place in PBL discussions lead to an increase in knowledge acquisition and recall. Furthermore, it has been proposed that students' cognitive development depends on the forms of language they encounter in their educational settings (Loftus, 2006; Visschers-Pleijer *et al*., 2006a). For example Rivard and Straw (Rivard and Straw, 2000) have looked at the way talking is influential in the development of scientific concepts in students. Of relevance to the methodology described here Chye (2006) has used discourse analysis to evaluate the instructional discourse of PBL students with the aid of transcripts of videotaped sessions.

A few studies have recorded, transcribed and analysed PBL discourses (Visschers-Pleijers *et al*., 2004; Visschers-Pleijers *et al*., 2006b; Hak and Maguire, 2000; van Boxtel *et al*., 2000). Using transcripts de Grave, Schmidt and Boshuizen studied the relationship between PBL tutorials and changes in cognitive processes (de Grave *et al*., 2001). Although these authors did use data from real PBL sessions, they created an experimental design in order to conduct the study and used stimulated recall from the participants to assess the changes in cognitive process.

Kamin *et al*. (2001) measured critical thinking in transcribed PBL discourses using qualitative research software (QSR NUD*IST) to develop themes based around deep and surface thinking, and Visschers-Pleijers *et al*. (2004) developed a coding scheme to measure communicative functions. Kamin *et al*. (2001) have even analysed on-line asynchronous PBL discourses. More recently, Yew and Schmidt (2009) analysed the verbal interactions taking place in a PBL group in order to increase understanding of the learning process taking place. They collected data from one cycle of PBL, corresponding to three hours of self-directed learning time and 4.5 hours of PBL sessions/meetings. Almost 1000 utterances were coded according the Van Boxtel code system (van Boxtel *et al*., 2000). This process of looking for themes and generating coding schemes is extremely powerful in capturing the richness of such data sets, however it is extremely time-consuming, error prone, it is subject to researcher bias and is not feasible to apply to larger numbers of PBL sessions or students interactions (Koschmann *et al*., 2000).

What is required are methodologies capable of making use of recent advances in computer-based text processing that can help researchers perform analytical functions, while providing a solid theoretical framework for data interpretation. Two electronic text analysis methodologies Corpus Analysis and Electronic Content Analysis provide a possible solution.

Corpus analysis, well established in the field of linguistic studies, is where the word corpus means a body of text (Adolphs *et al*., 2004). Corpus analysis can, in fact, lead to a more 'evidence-based' approach to the study of language in this particular setting (Adolphs *et al*., 2004) and many have argued the usefulness of using linguistic knowledge to improve text analysis (Roberts & Popping, 1996). An exploratory study using corpus analysis of one PBL cycle, reporting the use of this new methodology to explore and analyse the 'talk' that takes place during the PBL process. This study, represented a 'proof of concept' in that it demonstrated the possibility of enumerating and analysing student's technical vocabulary and their explanatory and reasoning skills (Da Silva & Reg Dennick, 2010). Based on nomological networks established within linguistics is possible to use language to “open the window” to some of the cognitive activities taking place during the PBL discussions. Corpus analysis will be the main methodology used in this study, allowing the gaining of insight on “how” are the students exploring and learning from the PBL case.

Electronic content analysis allows for automatic identification of main themes and sub-themes, allowing the researcher to do several types of search and query (Kondracki *et al*., 2002) . In the present research this methodology was used to gain further insight into the content being discussed in the sessions, that is understand “what” is being discussed, and supplement the (main) results from corpus analysis.

A.2. Corpus Analysis Definition and Description of Uses

A.2.1. Background and Definition

Background

Corpus Analysis is an electronic text analysis methodology used in corpus linguistics. Corpus linguistics is the field of linguistics concerned with the study of language use in the "real world". Corpus linguistics focuses on E-language, or external language use rather than I-language or the internal structure of language. The internal language approach has been more influential in studying linguistic field *space*, especially using the theories of Chomsky and other generative grammarians on the development of language and language structures (Melina, 1997). This approach characterises language as an error rich data source, and argues that it only possible to study it under very carefully designed laboratory settings, where one can control and minimise error. This also means that research has to be done using small sets of data that are explored in a very fine and detailed manner (Meyer, 2002). Corpus linguistics does not require experimental data sets and supports the use of natural occurring text, with only a minimal input from the researcher as the best path to understanding language and its structures. It deals with errors on the premise that, if the sample of data collected is large enough the language errors will not have a substantial impact on the conclusions. Also because it wishes to study language in the "real world" it argues that naturally occurring errors in language are a reflection of the richness of language and therefore they ought not to be eliminated but analysed (Leech, 2005). There are however, inside corpus linguistics, different views about how data should be analysed. Sinclair's works advocates minimal annotation[15] of corpora in order to let the text reveal its own meaning. On the other hand, authors like Quirk (1986) and others advocate that extensive annotation of texts should be used as it improves rigour and understanding of linguistics texts.

No matter what approach to annotation is used these authors support the core assumptions of corpus linguistics as providing the best approach to the study of language use, as opposed to the methods adopted by Chomsky and other generative grammarians.

15. Annotation is the process of adding additional information to the text. (Leech, 2005)

Another key difference between these two approaches lies in the adequacy level[16] they meet. Adequacy levels were defined by Chomsky and represent a taxonomy of linguistic theories, according to their methods and power of conclusions. There are three levels, the first being a descriptive/observational level and the last or highest being the explanatory. Theories at the first level describe reality, and it is here that corpus linguistics tends to be classified. At the highest-level theories can provide explanations and predict reality. This is the level at which generative grammarians classify their theories (Boeckx, & Hornstein, 2003). One can actually say that these represent two clear distinct paradigms of linguistic research (Leech, 1992). Discussion between both approaches and their authors has been extremely prolific, and is an important part of the history of linguistics as a research field. It is important to note that, since it is not so concerned with language origins or language theories that can predict language development, this allows corpus linguistics to look at language as a communication tool making this branch much more relevant and applicable to other fields (Meyer, 2002).

In the beginning of the nineteen-sixties two pioneering groups of researchers developed what we can now call the landmarks of modern corpus analysis. One was the work of Randolph Quirk (Quirk, 1996) and his colleagues at the Survey of English Usage group known as the London–Lund Corpus (LLC). This corpus consisted of the compilation of around 500,000 words from spoken sources, fully transcribed and compiled. Apart form the enormous endeavour involved in compiling such a large amount of data in the *pre-computer era,* its analysis pioneered the use of computation analysis in corpus linguistics, opening the doors of the field to the use of technology and leading to the birth of modern corpus linguistic (Leech, 2000).

The other contribution was the work by Kucera and Nelson at Brown University, US, published in 1967 with the title Standard Sample of Present-Day American English (the Brown Corpus). This corpus consisted of a collection of around 1 million words extracted from written texts published in 1961 in the U.S. The Brown Corpus collected data from a wide range of sources in order to represent the main current (at the time) uses of American English. This corpus also made use of complex computational analysis based on knowledge from several disciplines, which represented a novelty at the time (Leech, 1992; Leech, 2000).

16. Adequacy was defined by Chomsky as one of the key elements linguistic research should meet. (Chomsky 2005).

Both of the above mentioned publications were done in the pre-computer era, a time when research in corpus linguistic was extremely demanding and time consuming. Collection of large corpora were very difficult as accessing the materials was much more difficult then today, with fewer transcription options available, which were more time consuming. Then annotation, tagging and classification would have to be done manually by researchers. Due to the extent of the work involved this was a highly challenging and time consuming task.

Developments in technology meant that data collection, transcription and annotation of corpora became much more accessible to researchers. This meant that they could now make use of larger corpora that could be easily shared within the research community (Melina, 1997). This led to a much more prolific research field and to an increase in its support within the linguistic community. Leech (2000), in a paper about the outcomes of corpus-oriented research, mentions only eleven corpora of spoken language as some of the most useful to this research field, all of them containing large samples of uses of language with a combined total of over 30 million words from a wide variety of sources, contexts and countries.

Technology also poses new challenges, as the traditional methods of annotation and compilation of corpora had to be adapted and tested against this new reality. New methods and techniques of analysis emerged, along with the new software. Many of these tools were actually developed in very close collaboration with linguists, which helped computational linguistics move from being an area of computer science to become increasingly under the umbrella of linguistic departments (Adolphs, 2006; Melina, 1997)

Corpus linguistics' wider application to other fields of study also increased significantly and in particular there were applications to healthcare and healthcare education contexts. Corpus linguistics is nowadays a different and more agile research field than when it was first developed. Developments in technology and increases in the transcribed corpora available to researchers have had a major impact leading to what Leech (2000) described as the *corpus revolution* (Leech, 2000).

Definition: Corpus Analysis

C*orpus analysis* is the methodology that studies language (texts) from different sources (e.g. spoken, written) by classifying words in a text according to predefined linguistic categories and analysing relationships between them (grammatical and semantic) (Adolphs, 2006).

The process of corpus analysis follows the generic steps of the "scientific method" (Rayson & Garside, 2000) the first step is to start with a research question the researchers want to see answered. After that, data (text) has to be collected and prepared for analysis. In corpus analysis this is done by selecting, building and annotating a corpus or corpora. The analysis is then performed and information is extracted that will lead to answer(s) to the initial question(s). These phases are summarised by Rayson, into the five key steps of corpus analysis (Rayson, 2008).:

"1. Question: Devise a research question or model 2. Build: Corpus design and compilation 3. Annotate: Manual or automatic analysis of the corpus 4. Retrieve: Quantitative and qualitative analyses of the corpus 5. Interpret: Manual interpretation of the results or confirmation of the accuracy of the model" (Rayson, 2008 pp. 519-520).

This process can also be referred to as the 3A perspective: Annotation--Abstraction--Analysis (Rayson, 2008 p. 211) **(see** Figure 4.1**)** introduced by Wallis and Nelson (2001) (Wallis, 2001): Annotation consists in coding the text, that being the process of adding descriptive and/or analytic information to the text (Wallis, 2001). Abstraction consists in the integration of data in a model or a dataset. Analysis consists of searching the dataset in order to extract qualitative (such as utterances) or quantitative (statistical measures) information necessary to answer the research questions. For example, exploring frequency of common nouns or the use of the word "freedom" in political documents (Wallis, 2001).

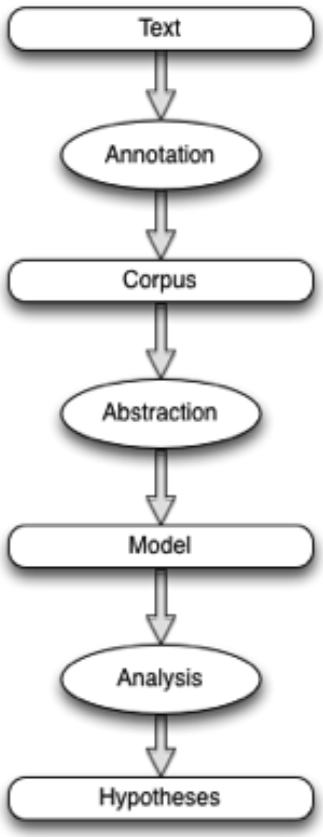
The above described process can be used within three main corpora research: Type I, II and III (Rayson, 2008).

The Type I approach focuses on the microscopic level of the text. This type of research looks into the smaller structures of the text, for example, words or specific grammatical structures. This is also the more "traditional" approach used by corpus linguists (Rayson, 2008). The Type II approach, also known as the macroscopic, focuses on the whole text or texts and not in its details. This approach focuses on variations and similarities across different texts for example in comparing grammatical structures across languages.

Finally, Rayson (2008) proposes a new type: type III. This is a combination of both type I and type II, and it can be defined a data-driven approach to corpora. Here the: *"decisions on which linguistic features are important or should be studied further are made on the basis of information extracted from the data itself; in other words, it is data-driven."* (Rayson, 2008 p. 521). With the aid of the new software tools available

in the post-computer era, this became a much more feasible and commonly adopted approach in corpus analysis. Type III, as the iterative inquiry method used in the systems analytics field (and recently action research (Goldkuhl 2008) is cyclical, and of spiral nature until a satisfactory level of analysis is reached. Here the enquire starts with an initial plan or question, followed by an initial analysis, results of that analysis are appraised and new aspects to be enquired are identified, leading to a new plan and the beginning of a new cycle. Each cycle will add more detail to the data analysis in a growing spiral way (Rayson, 2008).

Figure 4.1: The 3A perspective from text to hypothesis. Adapted from Wallis, S., 2001 pp. 312



This approach was adopted in our study of the PBL sessions. Two main reasons motivated this choice. First it was not our purpose to study the language characteristics from a linguistic point of view, but rather form a communication point of view. Second, because being the first time corpus analysis was being used in this context, the iterative inquiry approach allowed a continual redefining of aspects to search in the text, not possible with any of the other two "classical" approaches.

A holist impression of the text was needed (type II) in order to compare the texts produced on different moments and by different groups, based on frequencies of syntactic and semantic categories (e.g. common names, body and individual words). This analysis allows identification of the structures in the text that could better aid to answer the research questions and lead to a redefinition of the questions used to search the corpus. Following the first analysis, in order to extract data that would allow us to answer our research questions, a more detail analysis was needed. This in-depth analysis focused on specific syntactic and semantic structures (type I).

Text as Data

Corpora are a very special type of data. Texts are extremely rich datasets encompassing many variables such as the authors, the moment in time where it was produced and the context, among others. In order for a text to be incorporated into corpora and analysed using corpus analysis a few conditions have to be met as outlined below.

Corpus size determines the research possibilities, and it is usually determined by both practical and theoretical restrictions. A basic assumption of corpus linguistics is that corpora used in the analysis are large enough to be representative of a certain reality. Representativeness does not imply necessarily generalisability to a larger population/group then the one being studied, but an ability to represent the key elements reality to be studied. Therefore, a large enough corpus guarantees that sufficient recurrences of each lexical item are present to allow extraction of representative patterns. This ensures robustness of results and interpretations. If one aims to generalise language patterns across different contexts (e.g. different countries) or if one is interested in pure linguistic or lexical research then extremely large corpora must be used, as only those will be able to include key elements of such contexts (Adolphs, 2006).

Nevertheless smaller corpora can be equally useful if the aim of the study is not to generalise conclusions across a language but rather to understand a particular phenomena or reality via the study of language used. This is particularly true in applications of corpus linguistics to other areas, such as the ones the present study describes. Here analysis tends to make use of a mixed-methods approach by using a qualitative approach to aid data explorations. For example, if a French teacher uses corpus analysis to analyse in class discussion of a student group in order to identify common grammatical errors of those students in spoken French. In this situation it would not be enough to have one session corpus, this would be considered too small, the corpus should be able to provide a good representation of that particular group's discussions, for example sessions with different discussion topics. However, this would not be, neither aimed to be, generalisable to all students in French classes. For these purposes, smaller corpora carefully selected within the contexts of study are recommended. Corpus size should be always considered in relation to the other meanings without considering the other characteristics (described below) of the corpus and the study aims (Adolphs, 2006). The use of corpus analysis in the current research adopted a similar focus making use of small carefully selected corpora to understand a particular reality, not aiming at generalisations far beyond the research context. We collected and prepared a corpus of approximately 60 hours of first- and second-year medical students PBL discussion, this would be considered a small/medium size corpus. This analysis aimed at providing an understanding of how these groups of medical students approach and deal with the clinical cases within this context. Conclusions from this analysis were then used to add a different perspective to the discussion of the results of a quantitative cross-sectional study (presented in Chapters 7 and 8).

Representativeness is related to corpus size and is described as the characteristic that guarantees, in a sample corpus, all the possible characteristics of the reality reflected. Representativeness will be dependent on the research questions to answer and the study purpose. This characteristic will be ensured by the use of appropriate data collection procedures and sample methods. A simple example is the use of out-dated corpora, as it is not possible to use a corpus of spoken English from 1980 to identify characteristics of use of spoken English today as this would not provide representativeness (Adolphs, 2006). In the current research, making sure our collected corpus captured several discussions regarding different clinical cases, across

different modules, with different facilitators and in different moments in time ensured representativeness.

Another important characteristic in a corpus is balance. This is important if a corpus is constituted from a collection of different texts, such as studies on political ideology. These studies often use corpora encompassing many different types of text, from national television political speeches to political manifestos and laws approved by a certain political party. In these scenarios it is crucial to ensure that the balance of the texts within the corpus is maintained (usually according to their relevance for the analysis), otherwise valuable information contained in shorter, but valuable texts can be easily diluted (or even lost) when combining them into a single corpus (Adolphs, 2006). Balance of corpus was not an issue in the current research as our PBL spoken corpora contained exclusively transcriptions of PBL discussions.

One last characteristic is comparability, which guarantees that two or more corpora can be comparable. Corpora can only be compared if texts that are part of them are of the same type of data, meaning that the differences found between the corpora would be attributed to the variables under study and not to the differences in the type of material used within the corpora. An example is written text versus spoken text, where since grammatical and lexical structures are known to be different, any differences emerging in such comparisons will be due to this and not due to any other variables one wishes to explore.[17] These are simply different types of data, and as in any other research field data has to be of the same kind and from comparable samples, to be comparable. To ensure data is comparable one needs to guarantee appropriate sampling methods, data collection procedures and data preparation (to guarantee balance within the corpus) (Adolphs, 2006). In the current research comparisons where only made either between our PBL corpora and pre-established BNC (British National Corpus) of spoken speech or between our two PBL corpora, year 1 and year 2. At no moment our spoken corpora were compared with corpora built based on, for example, written materials (e.g. patient records, medical text books, facilitators guides) as those would not comply with corpus comparability requirements.

17. Fundamental research studies in corpus linguistics that aim to understand difference between spoken and written are an exception to this rule. Nevertheless there are specific methods to do this type of comparison and particularities to be observed (Adolphs, 2006). Those will not be described as they extend beyond the purposes of the present thesis.

A.2.2. Corpus annotation

Corpus Annotation in Corpus Analysis

In order for a corpus to be analysed it needs to be annotated (Leech, 2005; Rayson, 2008) by adding descriptive and/or analytic information to the text (see Figure 4.1). Annotation is key to corpus analysis as it is by this process that word characteristics are made explicit to the software for further analysis. This is what allows corpora to be searched and analysed using corpus analysis methods and it is also what guarantees that comparisons between corpora can be made (Leech, 2005; Adolphs, 2006). Different types of annotations can be combined to ensure that relevant information is added to the textual material according to the study purpose (Adolphs, 2006). Grammatical and semantic annotations are usually the first step in any annotation in corpus linguistics characterising each word based on syntactical and semantic value (Rayson, 2008). However other annotations are frequently added to these. A brief list of the most common types annotation used in corpus linguistics and its definitions is presented below:

POS (part-of-speech) tagging

PoS tagging or grammatical tagging is the most common form of annotation. This is the process of adding grammatical categories to a word or word combinations in the text such as noun, adjective, adverb etc. (Rayson, 2008).

Table 4.1: Examples of PoS tags used by Wmatrix2 (corpus analysis software used in this research)

CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)

Semantic tagging

Semantic tagging is the process of adding "meaning" or semantic value to words. An important part of this process is 'sense resolution', which describes the process of distinguishing the lexicographic senses of a word by combining an analysis of the POS tags and of the context surrounding a word (Rayson, 2008). This is the process by which each individual word is 'matched' against its possible dictionary definitions. For example, the word 'party' can mean a political group or an event involving a gathering of people. Underpinning corpus analysis software are extensive databases where many different dictionaries are combined in order to determine the semantic value of a word. However, often a word meaning depends on the context and it is fundamental to make sure those are differentiated, this is 'sense resolution' or disambiguation. Nowadays, due to the extensive research from the past on language use and software ability to deal with large databases, this can be done with few, if any intervention from human inspection. As the software's in-built formulas allow distinguishing between words meaning based on complex patterns identified by analysing association with other words, POS position and POS tag of the word. Using the example above, party as a political party will be frequently preceded by words such as 'member from a', ''belongs to' and associated with adjectives such as 'democratic', ' 'liberal', 'conservative'. The other possible use of the word party may be used in association with words such as 'birthday', 'Christmas', 'fun' or a time, geographical location, to cite just a few examples (Rayson, 2008). These elements make it possible for the software to perform the disambiguation process.

Table 4. 2: Examples of semantic tags used by Wmatrix2 (corpus analysis software used in this research)

A1	General And Abstract Terms
A1.1.1	General actions / making
A1.1.1-	Inaction
A1.1.2	Damaging and destroying
A1.1.2-	Fixing and mending
A1.2	Suitability
A1.2+	Suitable
A1.2-	Unsuitable
A1.3	Caution

A1.3+ Cautious

A1.3- No caution

A1.4 Chance, luck

A1.4+ Lucky

Grammatical Parsing

Parsing is seen as the process of analysing the text in order to understand its syntactic structure. Grammatical parsing is often a stage that follows from the POS tagging and attributes to sentences in the corpus a grammatical structure, usually based on the combinations of its words' POS tags, (Rayson, 2008).

Annotation of the PBL corpora

In the current research, we used CLAWS (Constituent Likelihood Automatic Word-tagging System) system of POS and the USAS (UCREL Semantic Analysis System) Sematic Tagging system to annotate the PBL corpus (Rayson, 2000). Corpus annotation was made automatically by the software used (Wmatrix2) using pre-existing dictionaries and a new dictionary created by the researcher to allow semantic tagging of medical technical language (e.g. names of drugs, bacteria, viruses). The procedure used to create this dictionary will be described in detail in Appendix 1.

CLAWS, the Constituent Likelihood Automatic Word-tagging System, is the part-of-speech tagging for English that is in development and constant progress since the 1980s in UCREL. This is the tagging system that attributes the grammatical tags, or POS tags (parts-of-speech tags) to words. The last version of this tag set, "Claws4", is used by Wmatrix2, to POS tag the 100 millions of words in the British National Corpora (BNC). The process of POS annotating the words follows the Word Class Tagging Guidelines as shown in Table 4. 3.

The accuracy of this tag set is remarkable, with an average of 96 to 97% accuracy depending on the type of texts.

For token[18] texts, Claws holds a 97% precision in spoken texts that is the extent to which incorrect tags are discarded from the final output. It also has 98.83% (for

18 Token in Linguistics is a " individual occurrence of a linguistic unit in speech or writing." Oxford Dictionary, 2012 .Oxford University Press [Available at: http://oxforddictionaries.com]

spoken texts) of recall that is the extent to which all correct tags are retrieved in the output tagger, even in the case when more than one category is attributed to a word (ambiguous tagging). This was calculated using a sample of 50,000 (45,000 written and 5,000 spoken) texts, those were tagged automatically by the system and manually checked by different users (human inspection) for reliability and consistency of the tags attributed[19] (Leech, Garside & Bryant, 1994; Rayson, 2000).

Table 4. 3: Wordclass Tagging schema for BNC2. (Adapted from Rayson, 2008 and UCREL,2000)

Process	Definition
Tokenization	Division of the text into each individual units (words and multi-expression word groups) *." (Rayson, 2000)*
Initial tag assignment	POS and Semantic tagging of every unit (word) in the text. Example *"word: e.g. words in -ness will normally be nouns." (Rayson, 2000 p. 2)*.
Tag selection (disambiguation)	Selection between different possible tags associated with the same unit (as above described using example of "party"). *"A method known as Viterbi alignment uses the probabilistic estimates available, both in terms of the tag-word associations and the sequential tag-tag likelihoods, to calculate the most likely path through the sequence of tag ambiguities. (The model employed is largely equivalent to a hidden Markov model.)"* (Rayson, 2000 p. 2).
Idiom tagging	Tagging is matched with a specific idiom template, depending on this some of the previous tags maybe corrected. "There are many cases in English where a sequence of orthographic words is best assigned a single tag. Such cases include so long as (a conjunction), and of course (an adverb)" (Rayson, 2000 p. 2).
Template Tagger	Template tagger is another programme within the Wmatrix2 software. This works as an elaborate 'searching and replacing tool'. It searches the tagging done previously for common, previously identified tagging error a and replaces those with correct tags which were hand coded into this tool by the developers. For example the word *after* in a particular context may not be a subordinated conjunction but a preposition, the template tagger has description of such cases and rule corrects many of those. It is also accessible to individuals using the software to upload and create their own classification rules to be used along with the previously inbuilt

[19] For more information and description of texts used please see: http://ucrel.lancs.ac.uk/bnc2/bnc2postag_manual.htm

	tagging systems (e.g. the medical dictionary created in this research) *" (Rayson, 2000 p. 4)*.
Post-processing: including Ambiguity tagging	"The post-processing phase has the task of producing output in the form in which the user is going to find it most usable." *(Rayson, 2000, p. 5)*. Different types of searchers are enabled at this point, different text output and queries are accounted for. Ambiguity tagging is also included for those cases where either the probability of each tag is similar, for example <NN1-AJ0> would be added as tag to a word that could be either a common noun (NN!) or an adjective. (AJ).

UCREL[20] Semantic Analysis System (USAS) is a classification system in which the semantic value of each word is organised hierarchically to allow text analysis and data retrieval. The USAS tagging systems was also used in several other projects to analyse spoken speech and discourse in interviews. From the USAS application to the Automatic Content Analysis of Spoken Discourse (ACASD) project and the Automatic Content Analysis of Interview Transcripts (ACAMRIT) project, improvements on the system have been introduced in order to better capture meaning in spoken textual materials.

"The semantic tags [now] show semantic fields which group together word senses that are related by virtue of their being connected at some level of generality with the same mental concept. The groups include not only synonyms and antonyms but also hyponyms. Currently, the lexicon contains nearly 37,000 words and the template list contains over 16,000 multi-word units." (Archer et al., 2002 p. 1). This vast lexicon of words and multi-words units is organised into a hierarchical multi-tier structure encompassing 232 category labels, representing fine-grained subdivisions of 21 high-level categories (figure 4.2 below). This classification system has a reported accuracy of 91% compared with manual tagging (Rayson, Berridge & Francis, 2004).

20. University Centre for Computer Corpus Research on Language (UCREL)

Figure 4. 2: USAS high-level categories

A general and abstract terms	**B** the body and the individual	**C** arts and crafts	**E** emotion
F food and farming	**G** government and public	**H** architecture, housing and the home	**I** money and commerce in industry
K entertainment, sports and games	**L** life and living things	**M** movement, location, travel and transport	**N** numbers and measurement
O substances, materials, objects and equipment	**P** education	**Q** language and communication	**S** social actions, states and processes
T Time	**W** world and environment	**X** psychological actions, states and processes	**Y** science and technology
Z names and grammar			

Although this is a comprehensive categorisation system, in the current research we a additional dictionary of medical terms was added to this software to improve accuracy of annotation (see section Data analysis below).

A.2.3. Abstraction and Analysis: Methods Of Retrieving And Interpreting Data In Corpus Analysis.

The first step in analysing a corpus is *"calculating basic information about the text or collection of texts"* (Adolphs, 2006 p. 48). Elements like total number of words, word lengths, number of paragraphs, number of tagged words and untagged, number of tags semantic and syntactical, are just a few of the most commonly items extracted at this stage. (Rayson, 2008; Adolphs, 2006). This descriptive stage is very important in collections of texts, and especially if the research involves comparison between corpora. There are two commonly used ways of retrieving and interpreting data in corpus analysis: frequency profiling and retrieving words in contexts (concordancing). These can be used individually or as combined ways to analyse the data.

Frequency lists

Frequency lists rank the words in the text according to the criteria defined by the researchers. Typical frequency lists used in corpus analysis are: word lists of single items or recurrent sequences, or keyword lists of single words or key sequences (Rayson, 2008; Adolphs, 2006).

Single word lists list the frequency of occurrence (relative or absolute) of words in a text or in different texts. They can be retrieved for all the words in the corpus, or only the more frequent words, according to syntactic or semantic tags or any other analytical tags, used in the text annotation. **Lists of recurrent sequences** are lists of words that co-occur frequently together in "clusters" (Scott, 1997) also referred to as ***collocations.*** These can be either common clusters because they are common expressions in a language, such as "I don't know" or "do you" or because they are common clusters found in a particular corpora, for example in our study: "mental health" or "learning disability".

This is a very useful tool to provide an overview of a text and it is usually a starting point for further analysis of the corpora. With the available corpus analysis software this can also be an iterative process of search in the corpora. For example the software used in this research (Wmatrix 2) allows the user to investigate the text starting from simple single word frequency lists. By clicking on the words it is possible to have access to the word in context (concordance), or by clicking on the tag it is possible to see the list of all words listed under that selected tag.

Frequency **lists of keywords** and **key sequences**, are lists of frequency that reflect not all the words in the text, but a particular word, group of words or sequence. They are extremely useful when the focus of the analysis is a particular aspect of the corpus and not the all corpus.

Concordancing is the act of retrieving words and their context from an annotated corpus. A concordance, also known as Key Words in Context (KWiC), is a piece of text in which a word or a cluster appears. Software allows us to choose the length of the context that is retrieved from the text. On some occasions it is useful to read all of the paragraphs in which a word appears, in others for example, a range of 120 words is enough, depending on the research questions to be answered (Rayson, 2000). A concordance facilitates the analysis of lexical patterns in the text, and provides meaningful contexts to words, facilitating disambiguation (if necessary) and interpretation of findings. As with frequency lists, concordances can be generated by single keywords, or by multi-word complexes (Rayson, 2008; Adolphs, 2006).

Statistical Measures In Corpus Analysis

Corpus analysis makes it possible to perform iterative inquires into the data, combining both quantitative and qualitative approaches (Rayson, 2008). In this process it is useful to use statistical measures in order to identify patterns, to understand relationships within the data, to compare different corpora or simply to better describe the corpus under analysis.

Typically corpus analysis uses data from frequency outputs (see above) as the raw data for further statistical analysis. Factor analysis, correspondence analysis or cluster analysis are the most common statistical methods used to research relationships within a corpus in corpus linguistics. These methods provide useful information to linguists about relationships between items (words, grammatical structures, others) in the corpus or they can be used to investigate the relationship between items in the corpus and other variables (metadata) such as demographic information (Rayson, 2002). In the present study the aim was not to understand the corpus from a linguistic point of view, but rather from a functional one. The aim was to identify the differences in a particular aspect of the text in order to understand and identify aspects, (such as questions, explanations, reasoning) defined as the early basis for the development of clinical reasoning. Log likelihood was the method more suitable for corpus comparison and identification of differences and therefore, the best method to answer the research questions of this study.

Finally it is important to notice, that the LL is a pairwise method, allowing only the comparison of two corpora simultaneously. This does not mean it is not possible to do multi corpus comparisons using LL, just that they have to be done in pairs, and that it is highly recommended to use control texts. These texts, usually subsets of a large corpus comparable to the corpus being studied (example: spoken English for analysis of oral conversations), and provide norm data to that subset (Rayson, 2002).

Box 4. 1: Log likelihood formula

$$LL=2*((a*\log(a/E1)) + (b*\log(b/E2)))$$

E1 and E2 are calculated according to: $E1 = c*(a + b)/(c + d)$, and $E2 = d*(a + b)/(c + d)$ and 'c' and 'd' correspond to the total number of words in the first and second corpus, respectively, and 'a' and 'b' are the observed numerical values (absolute

frequencies) of specific words or word categories in each of the transcripts. The LL is positive or negative according to the direction of the difference found. For significance level (see below) at P < 0.01, which is the common accepted value for significance, the critical LL value is ± 6.6 (Rayson, 2002).

Box 4. 2: Log Likelihood significance levels (Rayson, 2002)

95th percentile; 5% level; p < 0.05; critical value = 3.84
99th percentile; 1% level; p < 0.01; critical value = 6.63
99.9th percentile; 0.1% level; p < 0.001; critical value = 10.83
99.99th percentile; 0.01% level; p < 0.0001; critical value = 15.13

A.3.Corpus Analysis Application to Health Care

Research into language use in healthcare is not a new field. A simple search for *language + healthcare* using Google Scholar retrieves 684,000 hits[21], showing how prolific research has been. Initially focused on doctor-patient interactions, it has nowadays diversified into interactions between patients and other health professionals (nurses, physiotherapists, pharmacists, and others). However, as Adolphs has noticed most of these studies have *"small databases that have not originated from large data collections"* (Adolphs *et al*., 2004. p. 10).

Also most of these studies have been carried out by researchers outside the healthcare system, with only few involving healthcare professionals as part of the research team leading to a barrier to the application of research in healthcare practice (Adolphs, 2006, Elwyn *et al*., 2001). Another characteristic of many of the studies of language in healthcare is the use of discourse or conversation analysis (Roberts & Sarangi 2003) which *"strives to unpack the sequential orderliness of texts as active social phenomena as part of day-to-day institutional actions"* (Adolphs *et al*., 2004 p. 11). In general, the applications of corpus research in healthcare have been relatively underexplored, however, some advocated its use and shown its usefulness (Thomas & Wilson, 1996; Skelton *et al*., 1999, 2002; Skelton & Hobbs, 1999a,1999b; Crawford, 1999;

21. Search performed on Friday, 7th, October 2011

Adolphs *et al*., 2004; Brown *et al*., 2006; Seale *et al*., 2006, 2007; Harvey *et al*., 2007, 2008, 2012; Adolphs & Carter, 2007; Koteyko *et al*., 2008; Crawford *et al*., 2008, 2010; Schaufel *et al*., 2009; Friedman *et al*., 2009; Gooberman-Hill *et al*., 2009; Crawford & Brown, 2010; Pilnick *et al*., 2010).

All aforementioned advocate the use of corpus analysis as a methodology to explore communication in healthcare context; however, three particular publications have assumed this as their main focus representing those of Thomas and Wilson, Skelton and Hobbs and a book by *Brown, Crawford & Carter.* Both aimed specifically at demonstrating the value and potentialities of this methodology to analyse the complex phenomenon of medical and healthcare communication (Adolphs *et al*., 2004; Harvey, 2012). Thomas and Wilson interrogated a corpus of practitioner-patient encounters containing a total of 1.25 million words using an earlier version of the software also adopted in the present research (Wmatrix2) and concluded the results of using corpus analysis were comparable to "manual discourse analysis" (Thomas & Wilson, 1996 p. 92). More recently, in a book entitled '*Evidence-based health communication*' exploring several aspects of healthcare communication theory and research the authors defend *"Corpus linguistics perhaps offers the possibility for a more fully evidence-based approach to studying and learning about the uses of language in different settings."*(Brown *et al.*, 2006, p. 131). Two chapters (7 and 9) of this publication outline the application of corpus analysis to research, teaching and learning within the complex universe of communication in healthcare. Finally a more recent work from Crawford and Brown is of particular relevance for the current research. The authors not only advocate the benefits of using corpus analysis in healthcare research but also make clear the advantages of its use for education of healthcare professionals based on the long established experience of language education where this methodology is used in class-rooms to promote evidence-based education *"We argue that progress in the healthcare disciplines may be well served by taking a leaf out of the corpus linguists' book, and using a similar approach to deal with teaching and learning healthcare language. Moreover it might well be possible to link communicative styles, strategies and motifs to data concerning the effectiveness of healthcare interventions. In this way a more effective and evidence-based approach to healthcare language can be developed which will promote the best use of class time for trainees, and of scarce and expensive resources such as drugs and treatment facilities."* (Crawford & Brown, 2010 p. 21)

The work from John Skelton can also be identified as one of the most influential in using corpus analysis in this field. Skelton and Hobbs used corpus analysis as a methodology in three studies analysing a corpus of total of 373 primary-care consultations made by 40 doctors. A first study interrogated this corpus using mainly concordancing, type I microanalysis, to understand the use of medical jargon, the use of power in the consultations and how language was used "*to diminish the potential threat of the presenting disorder*"(Skelton & Hobbs, 1999a p. 253). In another study, this corpus was interrogated to understand the gender differences in consultation between male and female doctors. A combined approach, type III as described by Rayson (2002), was used, making use of means of words per consultation, relative frequency of "affective" and "facilitative" questions, and frequency and use of "mitigated directives[22]" (Skelton & Hobbs, 1999). Finally, a third study adopted the same approach (type III) using measures such as frequencies, association of words and concordances strings of first person pronouns such as 'I', 'we', 'me' and 'us'. The study conclusions allowed identification of some prototypical patterns of interactions and gain further understanding of the power relationships in these primary care consultations (Skelton *et al*., 2002). In another study, using a different corpus, of conversations between a haematologist and 61 seriously ill patients, Skelton and colleagues combined corpus analysis with protocol analysis to understand the relational aspects used by doctors and patients under these particular circumstances, mainly what roles do uncertainty, negotiated ambiguity and evaluative references have in the doctor-patient communication (Skelton *et al*., 1999). All these studies were published in prestigious medical journals, rather then those with a specific linguistic focus, showing the recognition of the usefulness of this methodology to the medical field.

A more recent study, methodologically similar to Skelton's work, used corpus analysis to understand practitioners' perspectives on compassion entitled "language of compassion". Two studies are presented in this paper. A preliminary study conducted with ten key nurse practitioners to ensure that it was possible to identify attributes of "compassion mentality" in the discourse of practitioners. Followed by the investigation of another corpus compiled of semi-structured interviews with 20 practitioners aimed exploring *"what each interviewee saw as the most important*

---

[22]. Directives are functional elements of speech that imply an order/direction to someone else actions. A mitigated direction is for example *"why don't you try …"* by opposition to what the authors define as an "aggravated directive" for example "try this…" (JSkelton & Hobbs, 1999a p. 218)

inhibitors and facilitators of compassionate care being provided in their local area" (Crawford *et al*., 2010 p. 9). The analysis reported focuses on four of these semi-structured interviews. This study allowed authors to draw conclusions regarding tensions in the healthcare culture (*production-line vs. compassion mentality*), prevailing paradigms and discourses among practitioners and highlight areas for further research (Crawford *et al*., 2010). This study is a good example of the potentialities of corpus analysis to research complex phenomena.

A note on the corpus size is important, as all the above studies used small/moderate size corpora, even Skelton's primary-care consultations corpus cannot be considered large if compared with those used in studies involving analysis of newspaper coverage (Seale *et al.*, 2007) or electronic communications, such as emails or online discussion forums (Harvey *et al.*, 2008, 2007; Harvey, 2012). However as mentioned before the size of the corpora must be aligned with the study aims and the extent of its conclusions.

Recently Adolph, Crawford, and Carter, along with colleagues from the University of Nottingham have formed *The Health Language Research Group* within the School of English Studies[23]. This group has been a pioneer in exploring the potential of corpus research in healthcare with two significant differences from previous research. First, its researchers are healthcare practitioners, or academics within healthcare sciences, as well as corpus linguists and academics from the linguistics field. Second, the aim of the research is to contribute to evidence for practice, by applying the tools of corpus analysis methodologies to qualitative and quantitative research. "*Combined qualitative and quantitative methodology drawing on tools traditionally used for corpus analysis can enhance our understanding of a particular health care setting*" (Adolphs *et al*., 2004 p. 12). For these reasons, and in light of the growing number of publications abovementioned, these authors have suggested that a new field could be created, within corpus linguistic called *Clinical Linguistics* (Adolphs *et al*., 2004). The members of this group have now a growing list of publications applying linguistics methodologies, mainly corpus analysis, to the study of a wide range of issues in healthcare communication.

Medical Corpora

23. http://www.nottingham.ac.uk/~aezweb/research/HLRG/index.html [accessed 10, September 2011]

Application of corpus linguistics tools to research healthcare settings may be recent, (compared with research into use of language in healthcare), but medical terminology has been a part of corpus linguistics for some decades now (Adolphs *et al*., 2004). Medical corpora can be subsets from a large corpus, or a specialised corpus. Medical subsets are frequent in a large corpus, such as British National Corpora (BNC). These categorise words that are frequently used and belong to the medical domain (McEnery & Daille, 1993; Habert *et al*., 2001). These are usually not very large or highly technical. Specialised medical corpora are usually based either on information from practice (narratives, medical records, discharge information or medical reports) or published material (reports and journal publications) (Habert *et al*., 2001).

The MENELAS project[24], is an example of the first type of corpus, based on patient discharge summaries, that aims to integrate this information with other hospital information systems in order to make retrieval of patient information easier for practitioners and to create a uniform European Hospital Information System. Examples of such corpora have been increasing in recent years, currently the national centre for text mining (NaCTEM) website[25] lists a total of ten different bio-medical corpora, most driven by compilation and annotation of words contained in specialist journal papers/abstracts or, as the BioMed Central's corpus compiled by the publishers themselves. These include general biomedical corpus (e.g. BionINFER[26]) and more specific such as the GENETAG[27] is corpus of 20,000 sentences with genes and proteins extracted from MEDLINE. This expansion is led by the need to create better more effective computer systems to help practitioners, scientists and researchers to search, retrieve and interpreted the growing bodies of biomedical information and has been headed by computer scientists or multidisciplinary teams. But some (e.g. the aforementioned research group in healthcare communication, Crawford & Brown, 2010), have started to apply these resources to healthcare education.

The present study will apply a XML medical lexicon developed (Nhàn *et al*., 2005) by the Linguist String Project (LSP). This project started in 1965 in New York University with the aim of facilitating retrieval of scientific information from large databases of

---

24.http://estime.spim.jussieu.fr/Menelas/menelas.html#SECTION00020000000000000000 [accessed 9 October, 2011]

[25]. http://www.nactem.ac.uk/resources.php [accessed last 05.02.2013]

[26] http://mars.cs.utu.fi/BioInfer/ [accessed last 05.02.2013]

[27] http://www.nactem.ac.uk/resources.php [accessed last 05.02.2013]

publications, such as PubMed, through improving Natural Language Processing (NLP). One of the achievements of almost 40 years of research has been a very complete medical knowledge lexicon (Nhàn, 2007)' that we have used to complement the medical subset of the BNC corpus that enabled us to carry out our analysis.

A.4. Electronic Content Analysis

Background, definition and applications

Electronic content analysis software has been developed based on research carried out into natural language processing, textual data mining, artificial intelligence and linguistics. As with corpus analysis software, automatic content analysis software has the ability to code the text automatically, based on previously defined categorisation methods derived from measures of statistical association and co-occurrence of words in large bodies of texts (Smith & Humphreys, 2006; Sowa, 2000). Also this software groups the codes, into larger units or 'themes', that are identified based statistical measures of frequent association in the text and on previously defined hierarchical networks of codes and by the co-occurrences of codes within the texts[28] (Smith, 2003; Smith & Humphreys, 2006). As in corpus analysis, the words are the basic units to be coded (or tagged) by the software. However, tags associated with each unit and the methods of extracting meaning from the text differ. While corpus analysis uses syntactic, grammatical and semantic pre-established dictionaries based on linguistic research to match the words with POS and semantic categories, electronic content analysis focuses on the semantic network of words, and uses this units to "bootstrap" a thesaurus of supporting terms to define that word (define a 'concept'). Those concepts are then allocated it to high order categories ('themes'[29]) based on mathematical clustering methods (Melina & Zuell, 1999; Melina, 1997; Smith, 2003; Stubbs, 1996). In electronic content analysis the concepts and 'themes' are *"derived and named solely based on co-occurrence in the texts"* (Hewett *et al*., 2009, p. 1735) while in corpus analysis the categories names are predefined. In summary the

[28] For more detail on measures of co-occurrence and statistical validation of the measures used in electronic content analysis is provided in (Smith, 2003; Smith & Humphreys 2006; Smith, 2000b; Smith, 2000a)

[29] We will use the word theme in inverted commas ('theme') to distinguish between the definition of theme in electronic content analysis and theme in classic content analysis. Themes in electronic content analysis are groups of salient concepts co-occurring in a text, these do not have the depth or human associated interpretations that themes may have in other fields of qualitative analysis have(Neuman 2003).

difference between these two types of software lies in the techniques used to explore the text and information sources used to code/tag the words and identify the themes or higher order categories (Melina, 1997).

This software helps the researcher in the time consuming and error prone task of allocating parts of the text to the codes, identifying of themes and higher order concepts in the text, improving the efficiency and reliability of the analysis. It guarantees that this process of coding is conducted with less human intervention making it less susceptible to research bias (Landauer & Laham, 1998). It also allows the identification of patterns that may not always be obvious to human inspection, especially if one is dealing with large bodies of text (Stubbs, 1996). Landauer & Laham have shown that latent semantic analysis, which is a type of electronic content analysis that is focused on improving marking and assessment, performs as well as humans even in something as complex as essay marking (Landauer & Laham, 1998).

"In English, a double negative forms a positive. (…) "However," he pointed out, "there is no language wherein a double positive can form a negative." A voice from the back of the room piped up, "Yeah, right" " (Paul Rayson, 2002, p. 86).

As the quote from Rayson above brilliantly exemplifies, language and communication are dynamic and '3 dimensional' with speech paste, tone, non-verbal communication and other elements adding to language and communication a layer of complexity that is not available by the simple analysis of textual material. For this reason the use of such software would not be advisable as the only method to analyse latent variables, as the software does not yet have the ability to analyse all the richness of human discourse or communication. It is useful when analysing explicit content and looking for explicit meaning in texts (Melina & Zuell, 1999). Responsibility lies with the researchers to ensure these methods are used within the scopes of their potentialities, and, most importantly their recognised limitations (Kondracki *et al*., 2002).

A.5. Advantages of Electronic Text Analysis Software

Technological advances made a *corpus revolution* possible (Leech, 2000). Several software options can be adopted to perform any of the above methods of corpus analysis. Some focus more on concordancing, others allow for both concordancing,

frequency lists or key word lists to be retrieved. The choice of software depends on the aspects that need to be analysed in order to provide answers for the research questions. Adolphs (2006 pp. 80-81), Melina (1997), and Kondracki *et al*., (2002) highlight some of the advantages of corpus analysis software:

Reduces the bias. Coding or tagging of the text is done automatically, using objectively defined computer rules, that can be defined by the researchers or be pre-existing categories, increasing the reliability and replicability of the classification accomplished.

Makes possible analysis of large bodies of text (data). For example, the latest versions of *Birmingham Bank of English* corpus have a total of 450 million word of present-day English use. Detailed analysis of such large amount of data would be extremely difficult without appropriate software (Kondracki *et al*., 2002).

Facilitates identification of patterns. Semantic or syntactic patterns are often too embedded in the text to be detectable by human inspection alone, whereas software is able to easily identify such patterns. A common example is the repetitive use of certain words or the choices made regarding synonymous words associated with certain political ideologies (Rayson, 2008 p. 8)

Allows easy manipulation of detail and type of analysis to be flexible. Software allows the researcher to choose the units of analysis accordingly to type of analysis (I, II or III) that better suits his/her study (Rayson, 2008). Additionally, the technology makes it possible to navigate from frequency lists to concordances and *vice versa*. For example, it is possible to view descriptive statistics about the text (numbers of words, types, frequencies) and just "click" to access concordances. The opposite path is equally possible, starting with a particular concordance or selected part of text, and obtain quantitative information about it. This flexibility is especially important when adopting a type III analysis strategy (Adolphs, 2006; Melina, 1997; Kondracki *et al*., 2002).

Facilitates collaboration. By allowing online storage, sharing and export of data into different formats, corpus analysis software makes it possible for the same data to be shared across researchers and research departments easily. This means on the one hand that it is easy for researchers to repeat and test studies done by others. On the other hand it means that aspects of data, not covered in a particular research, can be explored by others, making the analysis of a set of data much more complete (Melina, 1997 p. 2). Finally it is important to note that these methodologies are based on the verbalised content of the sessions therefore it is not possible to infer

about internal not verbalised processes. These methodologies are not able or aimed at capturing underlining cognitive processes that operate at the fast non-conscious level or those that the participants simple chose not to share with the group, that would require, for example, elaborated neuroimaging methodologies (Allen & Fong, 2008). However, it is possible to ensure that the situation in which data is collected stimulates verbalisation of some of the thought processes. This has been a grounding principle of psychological and linguistics research for many decades (Berry & Broadbent, 1984; Merrill, 2006). The characteristics of the PBL group discussions create the necessity for verbalisation, sharing of information and building of a shared explanation of the case (Hmelo-Silver & Barrows, 2008), this makes it possible to use methodologies such as corpus analysis and electronic content analysis.

B. Corpus Analysis: Validation Study

As was mentioned in the introduction to this thesis the aim of this research is to attempt to study the development of clinical reasoning. In order to do so it is important to understand the process by which the students start to approach clinical cases and how that process changes with increasing knowledge and skills. This understanding will inform the following phases of this research but it is a useful contribution towards a better understanding of this complex cognitive process crucial for the development of expertise.

PBL sessions are an ideal environment to observe this process. These sessions allow the researcher to observe how students interact with clinical cases, without the necessary disadvantages of artificially created experimental settings. Here the students are in their 'natural learning environment'.

Corpus analysis, the methodology from linguistics described in this chapter, allows us to explore the some of the cognitive activities[30] that students engage in during PBL sessions (Arppe *et al*., 2010), by an analysis of their verbalisations. Although many aspects could be explored, due to the aim of this research our analysis was focused on three particular types of utterances: questioning, explanations and reasoning, plus the frequency of technical terms used.

Questioning utterances were identified by either the question marker; explanations were identified by the "change and causality" semantic category; reasoning utterances were identified by subordinate conjunctions, a syntactic category that includes words such as 'because', 'if', 'then'. However after an exploratory analysis of the data using this method, one question is yet to be answered: *Does this approach have sufficient concurrent validity?*

This question is particularly important when corpus analysis is used on spoken text. Spoken text is embedded in a human communication process. Ironies are a classic example, if interpreted literally they might disclose the exact opposite of what the speakers means. Therefore one might refer to spoken speech as being "3 dimensional speech". Due to the complexity of analysing spoken speech, the answer to our question concerning validity becomes more relevant. Therefore we tested the ability

[30] This refers only to those cognitive activities that are accessible throughout verbalisation and can be inferred on that basis. A full discussion from the view point of different disciplines on the of the relationship between observable signs, such as speech, language and meaning can be found at (Osgood *et al.*, 1957). Further discussion on how computational analysis of text can be used to extract meaning of complex textual material such as metaphors can be found in *(Koller et al.*, 2008) and (Kintsch, 2000).

of 'clinical experts' to identify the semantic categories of questioning, explaining and reasoning in transcripts that had also been analysed by WMatrix2 (Rayson, 2000).

B.1. Methods

Sampling

The present validation study was carried out through classic content analysis[31] with a purposeful sampling procedure. A group of 20 practitioners taking part on the Master of Medical Education course at the University of Nottingham were asked to volunteer to take part in this study. Their task was to read one of two randomly selected and distributed transcripts from one first- or second-year PBL session, and highlight questioning, explaining and reasoning utterances.

Ethical considerations

All volunteers were explained about the study in a presentation done by the researcher, all were made aware their participation was entirely voluntary and all the data collected would be anonymous. All students in the room were given a healthy volunteer information sheet along with a consent form, the signed consent forms were keep by the researcher according to the rules of the University of Nottingham Ethics Committee.

The data collected was anonymously, the highlighted transcripts should not contain any personal information about the participants or any other element that would allow identification of the participants. Participants were made aware of this fact and asked, to comply with it by not signing or adding personal information to the transcripts.

The volunteers were given simple instructions, examples and a simple definition of each of the categories of utterances to search as shown in Table 4. 4. They were advised that in some cases utterances may contain more them one category and therefore they could allocate multiple code to them.

---

31. Classic content analysis is the method of classifying text accordingly to previously defined list of codes and themes. Hsieh, H.-F. & Shannon, S.E., 2005. Three approaches to qualitative content analysis. Qualitative health research, 15(9), pp. 1277-88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16204405 [Accessed July 18, 2011].

Table 4. 4: Instruction to participants (coders)

Colour	Type of utterance
Pink	Reasoning Any piece of text where the students are making connections, links, between facts or knowledge in order to reach a conclusion or/and decision. Reasoning may be linked with looking for probable causes or consequences. It may occur as suggestions: I guess If X then Y or as the examples below show it may be expressed as X may be due to Y or Z, but Z would do B not A …
	Example: Antibiotics have not been… in the long term they've not been successful. Maybe they were treating the wrong pathogen. Or maybe they didn't finish the course. .
Yellow	Explaining Any piece of text where the students are explaining something, a disease, biological mechanism, a body system etc. Example: If you have an acute allergic response to something in your airways so your airways swell you get bronchial constriction and you can air in and out of your airways…
Green	Questioning Any Piece of text where students are making questions.
	Example: Breathing difficulties dies and [unclear – 00:11:54] for exercise because you need to breathe harder with exercise don't you?

Data analysis and interpretation

After collecting the transcripts the selections marked up by the participants were then inserted into Nvivo8 by the researcher. Each participant was considered a different coder, given a Coder ID (userN) to keep the data anonymous and their selections inserted as coding choices.

The transcripts were also processed using WMatrix2 software to identify the questioning, reasoning and explaining utterances. As corpus analysis primary focuses on words, after the words in each of these categories were identified, a concordance of 120 words length was generated for each of the identified words. Each of these

concordances was inserted into Nvivo8 repeating the procedure followed for inserting the experts' selections, meaning the utterances identified through corpus analysis were marked as a selection made by a coder with a distinctive coder ID (W2), for easy comparison with the human coders.

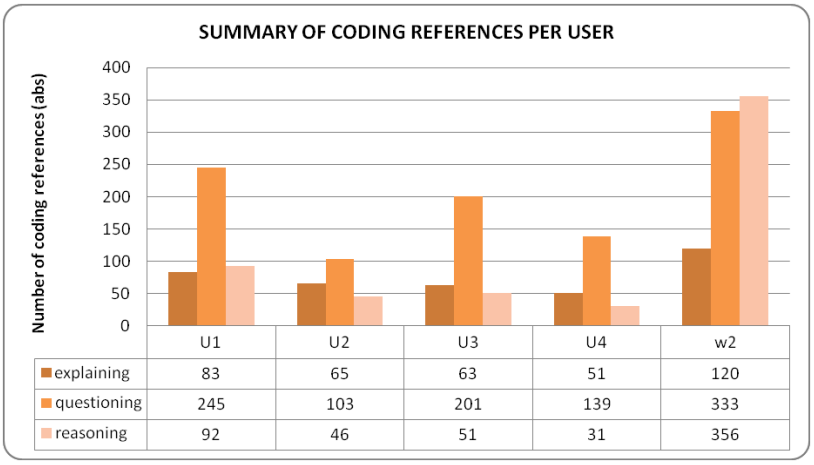
After all data were inserted into Nvivo8 values for agreement and Kappa values between groups (W2 vs. all userN) and within the users groups were calculated for each of the nodes (questioning, reasoning and explaining). Data from Nvivo was also exported to other electronic text analysis software, QDMiner (Provalis) and similarity index were calculated for the groups. This last index is automatically calculated based of degree of correspondence between the words coded within each category. Descriptive data, like the frequencies of nodes, of words per node, were also ascertained.

B.2. Results

From the 20 transcripts delivered 8 were returned with the coding completed. This is a clear limitation to this validation study and possible reasons will be outlined in the discussion section. A total of 3526 coding references were identified on both transcripts. All the users identified more questions than reasoning or explaining utterances (see Figure 4.3 and Figure 4.4). Reasoning utterances were the least identified utterances among users looking at the 1st year transcripts, whereas this was not found in the 2nd year transcripts.

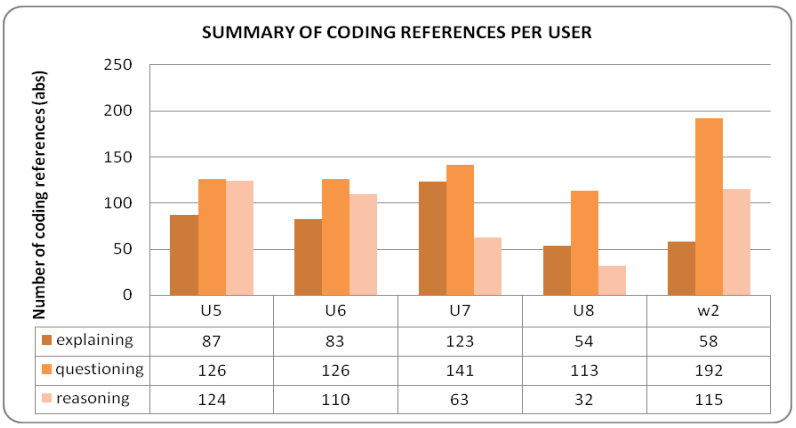
On both transcripts our corpus analysis software identified more relevant utterances than the users as can be seen in the figures below. This difference is particularly noticeable regarding the 1st year transcript coding selection for reasoning, where WMatrix2 identified more than 11 times as many examples than for user 4. In contrast, in the 2nd year transcript user5 identified more reasoning utterances than WMatrix2. At an individual level the frequencies of coding references for the 2nd year transcripts are much more consistent between users and between those and our analysis.

Figure 4. 3-Number of coding references per user 1st year transcript



	U1	U2	U3	U4	w2
explaining	83	65	63	51	120
questioning	245	103	201	139	333
reasoning	92	46	51	31	356

Figure 4. 4 - Number of coding references per user 2nd year transcript



	U5	U6	U7	U8	w2
explaining	87	83	123	54	58
questioning	126	126	141	113	192
reasoning	124	110	63	32	115

Agreement values are one of the most commonly used measures of inter-rater reliability. The agreement percentage equals the percentage of text selected by two coders, plus the percentage of text that has not been selected by any of them, minus the percentage of data that was only selected by one of them (disagreement).

Inter-rater Agreement (coder A and coder B) = Agreement between users - Disagreement between users. Agreement between users = percentage of text coded by A and B to the same category (A and B) + percentage of text not coded by A or B (Not A and Not B).

Disagreement between users = percentage of text only coded by A but not by B (A and Not B) + percentage of text only allocated to a category by B and Not by A (B and Not A). However this is a pairwise comparison, therefore in our study all the experts'

codings were considered as one group, by comparison with the corpus analysis software.

Agreement values may vary from 0 to 100%, values between 50% and 70% are considered good, and values between 70% and 90% are considered very good values. However, any value superior to 90% should cast doubts in the researchers' minds, as it is highly unlikely to be achieved without a strong bias in the research plan or procedures (Macqueen *et al*., 1996).

Our results show a very good agreement between users and our analysis with respect to questioning. Reasoning and explaining show medium to good values, interestingly slightly lower in the 2nd year transcript.

Table 4. 5:Agreement between W2 and all users

	1st year transcript	2nd year transcript
Explaining	69.78	55.03
Questioning	90.04	86.56
Reasoning	62.92	56.64

Nevertheless agreement values do not take into consideration the probability of coincidence, therefore Cohen's kappa coefficient was calculated.

"Kappa is a measure of the amount of agreement between two coders after statistically adjusting for agreement due to chance. Total agreement between two coders yields a kappa=1.00. Any disagreement produces a value <1.00, with lower values indicating larger discrepancies. Kappa takes negative values when there is less agreement than expected by chance alone (Fleiss, 1981)." (Carey et al., 1995 p. 5).

The Kappa values can vary from negative values, when agreement is less than expected by change, and 0.9. Good kappa values equal or are greater than 0.41 and very good for values equal or greater than 0.61. The Kappa index is a pairwise measure, not allowing for comparisons between more than two groups, therefore the results show in Table 4. 6 were calculated based on values shown in Table 4. 7.

In part, these results confirm the previous agreement results: questioning holds a substantial agreement both on average (0.62) and looking at individual comparisons in table (from 0.43 to 0.75). Explanations and reasoning utterances are not so consensual. These data also shows that the agreement values for *reasoning* are lower for the 1st year transcripts while *explaining* is higher in this transcript when compared with the 2nd year transcript. However, while the agreement values for *reasoning* and

explaining found previously are considered good, kappa values are lower then 0.20, therefore considered as poor agreement.

Table 4. 6: Cohen's Kappa values for agreement between W2 and nodes (general)

W2vs all users	1st year transcript	2nd year transcript
Explaining	0.26	0.19
Questioning	0.57	0.62
Reasoning	0.09	0.14

This apparent decrease in agreement between Wmatrix2 and the volunteers in our study is explained by the *chance-correction* introduced part of Kappa calculation. The debatable nature of chance-correction measure might be a possible reason for this result.

Chance-correction in Kappa calculation is a standard measure of agreement expected by chance for two users. This measure does not take in consideration the type of study, or the tasks given to the participants (Uebersax, 2010). *Chance-correction* calculation would be the same for two people flipping a coin, or as in this study, users selecting pieces of text according to instructions. Therefore some (Uebersax, 2010; Agresti, 1992) have argued this should not be calculated as a single measure independent of the study design and aims, but rather by statistical modelling according to the variables being studied (Uebersax, 2010; Agresti, 1992).

The poor results demand further inspection of each participant individual choice. Table 4. 7 presents the observed kappa values for agreement between Wmatrix2 and each of the categories. Results of questioning are consistent with previous results, being higher than on the other two categories. Nevertheless, it is possible to observe for all categories there is a substantial degree of variability on the coincidence between the experts and W2. For example for reasoning utterances, user 5 does not agree at all with Wmatrix2 (k=0.01) while for the same transcript user 8 reveals a moderated/low level of agreement (k=0.28). This variability might impact the overall results shown in Table 4. 6 as all the participants contributions had to be combined in order to be compared with Wmatrix2 (kappa is a pairwise measure)., therefore Table 4. 7 shows individual pairwise comparisons between Wmatrox2 and each individual user.

Table 4. 7: Cohen's Kappa values for agreement between WMatrix2 and each of the categories

	Wmatrix2 vs.	Questioning	Explaining	Reasoning
1st year	User1	0.71	0.25	0.17
	User2	0.44	0.28	0.09
	User3	0.75	0.21	0.09
	User4	0.43	0.31	0.10
2nd year	User5	0.62	0.17	0.01
	User6	0.67	0.04	0.11
	User7	0.61	0.22	0.16
	User8	0.57	0.21	0.28

Table 4. 8, Table 4. 9 and Table 4. 10 show the Kappa values for agreement between experts alone. Interestingly the values in both Table 4. 9 and Table 4. 10 are much lower than expected, especially the ones regarding to reasoning utterances identified. For example, for reasoning utterances very poor agreement is verified between users 1, 2 , 3 and 4 (1st year transcript) with all the kappa values (k<0.10) being lower then the overall agreement between all experts and Wmatrix2 (k=0.19). Similarly user 5 and user 8 also report a very poor agreement (k=0.10) in reasoning utterances in the second-year transcript. The low agreement between experts, might be part of the reason why the low agreement between all experts and Wmatrix2 was found; however, other reasons might explain these results, those will be discussed in the next section.

Table 4. 8: Questioning

	1st year			2nd year			
Kappa	User1	User2	User3	Kappa	User5	User6	User7
User1				User 5			
User2	0.48			User 6	0.64		
User3	0.71	0.50		User 7	0.60	0.70	
User4	0.43	0.40	0.47	User 8	0.58	0.56	0.55

Table 4. 9: Explaining

	1 st year			2nd year			
Kappa	User1	User2	User3	Kappa	User5	User6	User7
User1				User5			
User2	0.37			User6	0.48		
User3	0.37	0.54		User7	0.40	0.47	
User4	0.34	0.38	0.38	User8	0.37	0.50	0.48

Table 4. 10: Reasoning

	1 st year			2nd year			
Kappa	User1	User2	User3	Kappa	User5	User6	User7
User1				User5			
User2	0.06			User6	0.38		
User3	0.09	0.24		User7	0.20	0.30	
User4	0.10	0.28	0.25	User8	0.10	0.34	0.30

Figure 4. 5: Similarity index between nodes

*Similarity index is was automatically calculated using all the utterances code by the users and Wmatrx2 within each of the selected categories. The software (QDAMiner) aggregates all utterances per category and determines how similar the categories are based upon similar and co-occurring words between the words in each category. A similarly of 0.0 represents two texts with no shared words, and a similarity of 1.0 represents two identical texts.

Figure 4. 5 shows the similarity index for both the categories and the selection made by the different coders. Similarity indices describe how similar the utterances are selected in each of the groups, based on the type of words and syntactic structures.

Explaining and reasoning utterances are much more similar than questioning utterances. This might explain why the variability between users and between those and our analysis is much higher in these nodes. The fact that questioning utterances are distinct from reasoning and explaining utterances might have facilitated their identification in the text leading to a better agreement between experts and Wmatrix2 analysis. This will be further discussed below.

B.3. Discussion and Conclusion

B.3.1 Limitations of the Validation Study

Lengths of segments within codes

No limit was given to the coders on relation to the size (word count) of the segments of text to be coded within a single code. The only information given was "Any piece of text". This led to some differences within the coders. Although not major discrepancies were noted, it was noted that some participants tendency to highlight more extensively full phrases and/or paragraphs (e.g. user 8 reasoning code) while others were more restrictive in the length of the coding references increasing the number of such references used within the same text segment (e.g. user 1 questioning code). This fact made it very difficult to extract any meaningful conclusion only from the comparison of the number of coding references and may have had a slight effect on the agreement and kappa values. Although after the analysis this was identified as a limitation, and something to be done if conducting a similar study in the future, the differences were only subtle

Transcripts are hard to read

Only 40% of the transcripts were returned and coded by the participants even though they expressed initial enthusiasm for the task. We believe this might indicate that the task given was probably more difficult or more time consuming than the participants

anticipated. In fact, the transcripts of the PBL sessions are a different and unusual type of text for people to engage with (see in the Figure 4. 6 (below). Created by eight people working and discussing a case over a 2-hour period ideas flow, and speech overlaps, making it a very peculiar and perhaps difficult reading material.

Coders were not trained

Training the coders, usually by means of a group meeting where people are invited to start coding the text and discuss possible questions and differences in codes interpretation, is a common procedure in classic content analysis (Mayring, 2004; Weber, 1990).

The aim of the validation process was to compare the Wamtrix2 classification with the coding done by human coders with a healthcare background and interest and involvement in medical education. This purposeful sampling method was used to guarantee that the coders were a group of practitioners intellectually engaged with educational literature and research. This would be expected to, on one hand, increase homogeneity within the sample and, on the other hand, make the validation more meaningful, as these participants are more likely to be carrying out research in medical education than a generic pool of healthcare professionals. If nothing else they will be expected to conduct a year of research to complete their masters.

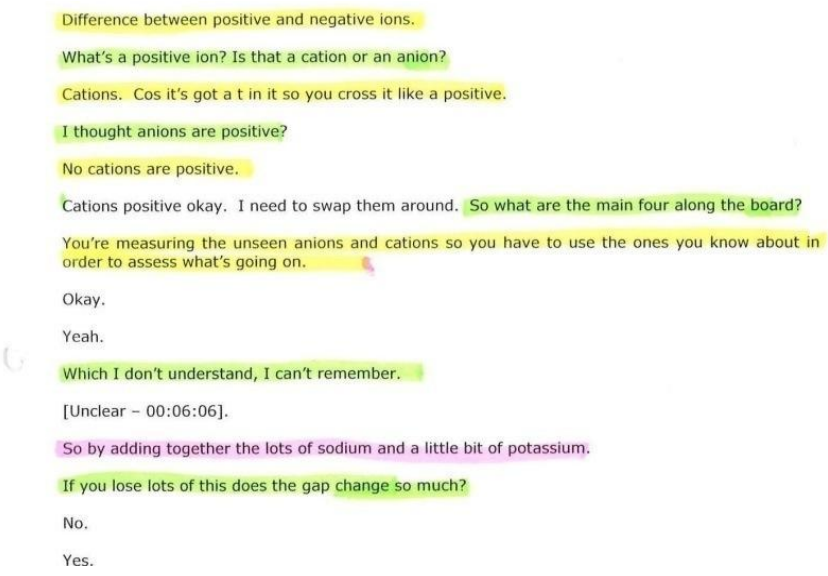
On this assumption we chose not to train the coders. As training the coders in our present study could have led to 'hiding' potential differences in interpretations or views reading what a 'reasoning" and/or 'explanation' would look like in the transcripts. Also it could have skewed participants towards a specific coding criteria defined by the researcher, which would decrease the value of their participation. However, instructions were given to the participants' along with the coding task (as shown in Table 4. 4). This an important limitation of the study, as large disparities between coders were found. That may well be because, although familiar with medical education literature and research, participants views on what constitutes a reasoning, an explanation or even (less) a question differ, and specially how those elements "look like' in a PBL transcript, differs more then we initially assumed.

Small number of coders per transcript and more transcripts

Also in face of this variability it would have been good to have a large sample of coders per transcript and possibly more transcripts. Unfortunately, return rate of coded transcripts was much lower then initially excepted, from the 20 transcripts

distributed only 8 were returned, this could have had an impact on the data of the validation study, perhaps if more had been returned it would be possible to identify patterns of convergence between coders that otherwise with only four coders per transcript were not identifiable.

Figure 4. 6: Scanned image of a coded transcript


Difference between positive and negative ions.

What's a positive ion? Is that a cation or an anion?

Cations. Cos it's got a t in it so you cross it like a positive.

I thought anions are positive?

No cations are positive.

Cations positive okay. I need to swap them around. So what are the main four along the board?

You're measuring the unseen anions and cations so you have to use the ones you know about in order to assess what's going on.

Okay.

Yeah.

Which I don't understand, I can't remember.

[Unclear – 00:06:06].

So by adding together the lots of sodium and a little bit of potassium.

If you lose lots of this does the gap change so much?

No.

Yes.

B.3.2. Conclusions

Questions are consensual and clearly different from other nodes

Questions seem to be the most easily identified nodes in this study showing good to moderate values of both of agreement and kappa index among experts and between users and Wmatrix2 analysis. Questioning utterances identified are more distinct than reasoning and explaining utterances. The fact that questioning tends to be marked by punctuation, that is question marks, might explain this result. Also this fact decreases the variability of its definition among users. It is a consensual fact among all English speakers that a question mark indicates a questioning utterance, although there are no agreed markers when it comes to explaining and reasoning utterances.

Reasoning and Explaining utterances are very closely related

The similarity between reasoning and explaining utterances indicates that there is overlapping between such utterances. Some phrases may simultaneously express

explanations and reasoning, leading some users to select them as reasoning and others as explaining or even as both depending on their own definitions. For this reason, and the lower agreement values observed in these two categories, in our subsequent analysis of the PBL sessions these were combined in a single category.

Reasoning and Explaining utterances are very difficult to identify intuitively by experts

The difficulty in finding objective markers for reasoning and explaining can explain the differences found both between users and between users and WMatrix2. The lack of such markers makes the coding more dependent on the expert's own views and definitions of the complex processes of reasoning and explaining.

Corpus analysis provides an approach to identify Reasoning and Explaining in transcripts

Corpus analysis uses an extensive network of syntactic and semantic relationships based on the analysis of extremely large bodies of spoken text collected during many decades (British National Spoken Corpus) to identify its categories and its "markers". In this case markers are words or sequences of words, such as '*because', 'if', 'then', 'consequently' or 'like'*, that are frequently associated with reasoning or explaining.

The use of corpus analysis to train teachers and facilitators to identify explanations and reasoning may be beneficial

One of the aims of PBL is to stimulate the students' reasoning and explaining ability. If these concepts are not well defined and consensual among facilitators it will be difficult to give appropriate feedback to the students regarding them. Similarly to what is done in the language education field, where corpus analysis is used to feed back to classroom teachers and groups (Adolphs, 2006), maybe also in PBL corpus analysis could be used as a basis for training facilitators to identify markers of reasoning and explanation during PBL sessions. This could help increase facilitators' awareness of these during the discussions.

C. Analysis PBL Sessions: Methodology

This section describes the methodology used in study 1: analysis of the PBL sessions, and how we applied aforementioned corpus analysis to the aims and questions of the present research.

C.1 Philosophical Approach of This Study

This particular study (study 1) is situated within the pragmatic approach. This paradigm, described in detail in the introduction chapter, provides the philosophical, ontological, epistemological framework that underpins all parts of the current research.

C.2 Aims of The Study and Research Questions

The research questions guide all the research process, they justify the choices made regarding methodology and the type of analysis performed and provides the basis for interpretation and discussion of results. Finally they also provide the framework that integrates this particular study into the overall research presented in this thesis. Below we present the aims, research questions and sub questions that emerged from the main research question and guided the study now described.

Aim 1: Understand the use of technical language in exploration of the clinical cases

Questions 1: Does the student's use of technical vocabulary (medical lexicon) change over time within year (longitudinal)? And between years (cross-sectional comparison)?

> Question 1.1 What are the differences in frequency of technical terms across the sessions/cycle/years of study?
>
> Question 1.2 What are the differences in the type and nature of technical terms across the sessions/cycle/years of study?
>
> Question 1.3 Is there any significant correlation between the clinical content of the case and the use of technical lexicon? Does that vary between the two years studied?

Aim 2: Understand the strategies used by students with little medical knowledge to discuss clinical cases.

Question 2. Can questions, explanations and reasoning taking place during PBL discussions contribute towards understanding of development of the early basis of clinical reasoning?

Question 2.1 Are there any differences in the frequency questions, reasoning and explanations between sessions/cases/modules/years?

Question 2.2 Does students' questioning, explanation and reasoning frequency differ based on the phase of the clinical cycle[32] they are exploring? Does it change between years (1st years vs. 2nd year)?

Question 2.3 Do students of different years use different strategies to approach clinical cases as measured by questions, explanations and verbalised reasoning?

By answering these questions we will contribute towards improving the understanding of the cognitive activities involved in PBL discussions and, more importantly, contribute to understand the process by which medical students, since their early days in medical school, start to develop clinical reasoning.

C.3 Study Design

Although relying on qualitative data, corpus analysis and electronic content analysis can make use of both qualitative and quantitative approaches to data analysis. Frequency listings and statistical measures such as the log likelihood are typical of the quantitative approach with concordancing and keywords in context more associated with a qualitative approach. The choice of using one or the other or both approaches is guided by the research questions posed.

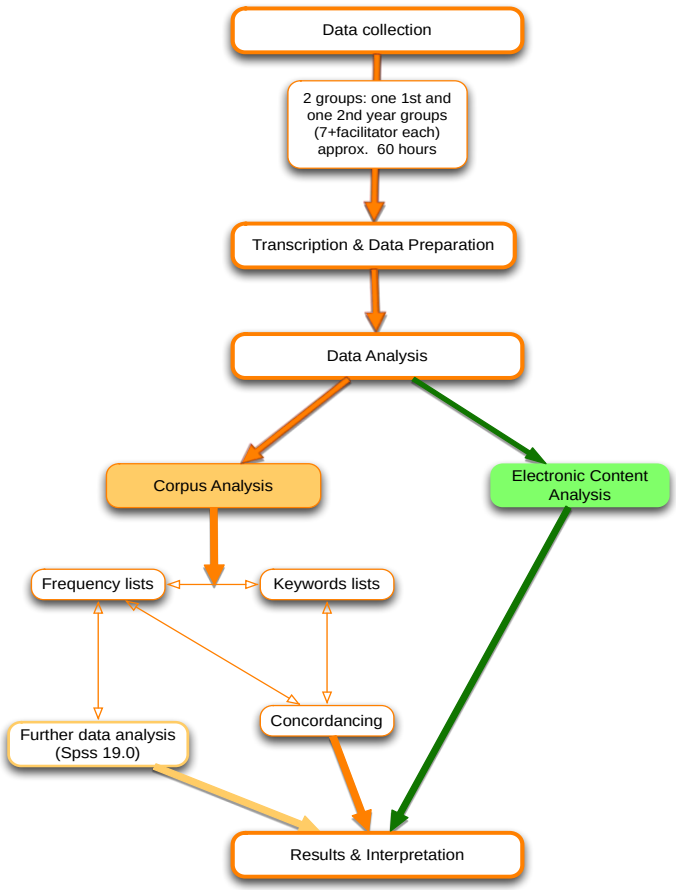
In the present study, analysis of PBL sessions, a mixed-methods approach was adopted (or type III approach described by Rayson, 2003). We combined both macro level analyses (frequencies and quantitative indicators) with a further exploration of particular utterances (micro level) in order to gain further understanding of the data and its context. This approach is in accordance with what Tashakkori & Teddlie (2003) define as a "truly mixed-methods approach" where iterative and complementary analyses from different perspectives and view points to take place in all phases of the study. Although a large percentage of the results are expressed in quantitative terms, during the analysis (data search and retrieval) concordancing was

32. Identified by the facilitators guides as follows: 1: patient presentation; session 2: exams and clinical findings; session 3: diagnosis and treatment.

frequently used. This method of corpus analysis was used to verify classifications done automatically by the software, to understand particular frequencies, co-occurrences, or collocations and whenever interpretation was required or disambiguation of utterances or words.

The present study is a cross-sectional study with intra and inter-group comparisons, conducted with two groups of students in the Graduate Entrance Medical School of University of Nottingham. Two groups selected were followed during four months of the academic year 2008/2009. During this period all the PBL sessions of both groups were video and audio recorded in a total of 26 sessions of first-year groups and 15 sessions of second-year groups, giving a combined total of approximately 60 hours of recordings.

Figure 4. 7: Schematic/outline view of the study design



In order to answer the previous research question the present study made use of corpus analysis adapted to the study of PBL sessions, as described in the previous section. Additionally electronic content analysis of the sessions was performed in order to identify the main themes (clinical content) emerging. These helped to redefine the data search within the corpus analysis software and the final data interpretation and discussion. Electronic content analysis was used a complementary method of data analysis in order to understand "*what (content)"* was being discussed during the PBL sessions, while corpus analysis was used to understand "*how (process)"* was it being discussed *"what were the strategies involved".*

C.4 Sampling

Sampling procedures must reflect the nature and purpose of the study (Neuman, 2003; Merriam & Sharan, 2009). In this particular case, our aim was not to generalise results, but rather to gain insight into the process by which, early in their courses, medical students approach, discuss and reason through clinical problems provided by the clinical scenarios of the GEM course. Consequently as it was not theoretically sensible to use probabilistic sampling methods, therefore a non-probabilistic convenient sampling method was used.

All the students and facilitators of both cohorts of University of Nottingham Graduate Entry Medical (GEM) School were invited, individually, to take part in the present study. After collecting individual agreements from students and facilitators data was aggregated allowing the identification of only two groups in which all elements had consented to take part in the study. Although recognising the limitation of this method, as compared, for example, with purposeful or a judgment sampling, where the researcher selects the sample based on important characteristics (Neuman, 2003), this method is suitable for the current research as all participants comply with the minimum requirements of the study, that is being part of a medical degree adopting a PBL curriculum where basic sciences are learned through frequent planned discussions of clinical scenarios.

C.5 Ethical Considerations

Informed consent and voluntary participation

Ethical approval from University of Nottingham Medical School Ethics committee was obtained prior to any data collection procedures. At the start of this research the PBL groups were already defined. This is a complex process carried out by the GEM school educationalist in order to ensure the groups are balanced and have the necessary human characteristics to allow them to be functional. Therefore it would not be sensible or ethical to re-arrange them just for purposes of this research. Consequently in order to be able to conduct the study, we needed all the students and a facilitator in, at least, one PBL group per year to volunteer.

To avoid any type of peer-pressure, and on advice of the ethics committee, all the 1st and 2nd year students of the GEM programme were invited to participate in the study on an individual basis. An announcement during a lecture to explain the study and purposes was made to both years, and was also given during two (one of 1st year and one of 2nd year) facilitators' meetings. During this moment all individuals were given the study information sheets and consent forms, after collecting the individual consent forms the researcher matched these against the groups' composition and facilitators. Although most facilitators consented or expressed consent conditional to the groups will, only one group per year (total of two groups) were identified where all members had, individually, agreed to take part. Previously to the change of facilitator the researcher confirmed the new facilitators had given previous consent, and contacted them individually to ensure they were still wiling to take part. All participants were informed in several occasions that they would have the right to withdraw from the study at any point without any required explanation.

Confidentiality and Anonymity

Video and audio files are very sensitive materials, as they contained a high level of personal and identifiable information. All video material, where individuals could be identified was kept strictly confidential with only the researcher and the supervisor of the researcher having access to it. These were not to be used in any type of conference or educational materials without previously seeking new individual consent from participants for its use. These video recordings were used only by the researcher (transcripts were made based on audio material only) to crosscheck, correct and clean transcripts (e.g. delete facilitators contributions). These were kept

in a secure location to which only the researcher and the supervisor had access within the Medical education unit.

Audio files were kept under the same secure conditions as the video files. Audio recording of the session were sent for professional transcription in accordance with previously granted consent by the participants. As the aim was not to look at individual contributions but the group, no identification of speakers in the transcript was made. The researcher reviewed individual transcripts to remove any identifiable information (e.g. such as names)[33]

Additionally the students were given the responsibility to manage the recordings and were given clear instructions to stop video and audio recording if the discussion involved any confidential or particular sensitive aspects, this was key to safeguard the anonymity and interests of all participating in the study. Also at no moment any identifiable material was shared with anyone involved in assessment or evaluation of the students or facilitators.

After the cleaning stage, in which the aid of video material was required to review particular aspects of the transcripts (e.g. remove facilitators contributions), all the data analyses were performed on anonymised material as were all the results of the current study.

Respect for the learning environment

The respect for the learning environment is one of the most important aspects of any educational research (Berliner, 2002). When research takes place in real situations, rather than artificially created situations it is the researchers responsibility to ensure the research does not undermine any part of the educational experience. Research, such as the one here presented always has an impact on the learning environment (Berliner, 2002; Cobb *et al.*, 2003). However, it is the researchers obligation to ensure data collection methods are not likely to have a negative impact or destabilise students learning in any way.

C.6 Methods Of Data Collection

Data collection took place during four and half months of the academic year 2008/2009, from October to February. During this period all the PBL session of both

[33] This process is described in more detail in the Data Analysis section.

groups were video and audio recorded in a total of 26 sessions of first-year groups and 15 sessions of second-year groups, giving a combined total of approximately 60 hours of recordings.

A crucial aspect is to avoid the "researchers'/research effect", that can be defined as the impact that the data collection, test or observation, and researcher can have on the behaviour of the participants, changing their normal behaviour and creating a confounding effect (Trochim, 2006). Eliminating this effect is, of course, impossible; however, it can be minimised and as can/should be the interference of the researcher into the articular dynamics of the PBL groups. Therefore, it was decided to allow the groups to record their own sessions without the presence of the observer/researcher. The researcher was responsible for setting up the equipment before each session and collecting it in the end, creating an opportunity to talk with the groups and understand if any problems had been experienced during the recording. These were kept in the researcher's audit diary.

This not only contributed to the authenticity of the sessions, but also most importantly was aligned with the ethos of the PBL curriculum and group rules. These groups enjoy great autonomy, and the responsibility for the management of the group and the group tasks lies with the students not the school or the facilitators, therefore also, as research participants, they should be the ones to decide what they agreed to be recorded and what were discussions about the group dynamic or sensitive issues that they, as a group, would prefer not to be seen by the researcher. Both groups agreed that they would like to have control over the recordings, so that, the recording would only capture case-based discussions and not any moments where irrelevant or personal issues were discussed. Following this introduction to the study, a first session was recorded in order to allow the groups to familiarise themselves with having the recording equipment in the rooms, and to test for possible equipment problems. These sessions were not considered for data analysis purposes.

Another critical aspect is to ensure the quality of the recordings for transcription. This was guaranteed by using, when necessary, software to improve the quality of the sound files and to eliminate background noise. Correct placing of the recording devices in the rooms also aid in ensuring this quality. Finally the use multiple recording devices, (video recorder and audio recorder) placed in the PBL rooms also allowed for the researcher to correct the transcripts done based on audio files by the professional transcriber.

C.7 Methods Of Data Analysis

Data preparation

After data is collected, it needs to be prepared for analysis. **Data Preparation** included: Transcription, Transcript "cleaning" and Pre-software annotation.

Transcription

The researcher must decide what parts of the recordings are to be transcribed, and what type of transcription is to be made, according to the research question to be answered. The three main types of transcriptions are: semi-transcription, full transcription and colloquial (Taylor & Gibbs, 2010). Semi-transcription, is a type of transcription where only the more relevant parts are transcribed exactly as they were said (*verbatim*), information is summarised into clearer or shorter sentences, and often notes by the researcher are made in the transcript.

Full transcription, describes the type of transcription in which all that was said is included in the transcript. In this type of transcription the person making the transcript will *"tidy"* the transcript, meaning it will replace, for example parts of words by the full word, or ignore irrelevant sounds that do not constitute words. When the speaker says "kids wi' som'd'y" it is transcribed as "kids with somebody" (Taylor & Gibbs, 2010). Colloquial transcription is the type of transcription where colloquial expressions, semi-words and sounds are included in the transcription. For example "was nae good." (was no good) (Taylor & Gibbs, 2010).

Semi or colloquial transcriptions cannot be used with corpus analysis. Using these types of transcription would make automatic software annotation extremely difficult and would cause unnecessary errors to occur. Corpus analysis requires all the speech to be transcribed with exact transcriptions of words used and the exact sequences of the discourse in which they were used.

Transcript *"cleaning"*

By transcript *"cleaning"* we mean the phase of data preparation in which the researcher deletes from the transcripts any irrelevant material, for the purposes of the study, and reviews them to guarantee the necessary quality in order to proceed with the analysis (Taylor & Gibbs, 2010). This was done manually by the researcher with the aid of the video files. Each session transcript was reviewed against the video recording. During this process, mistakes and errors in the transcript were corrected,

irrelevant material deleted and the facilitators' contributions were deleted. Due to the nature of PBL sessions, with eight people discussing a clinical case, often side conversations can take place. These moments were indicated in the transcript notation, between square brackets, allowing the researcher, when reviewing the transcripts to decide on whether or not to include these. For the present study side conversations not related with the case or learning (e.g. conversations about the football results the night before) were considered not relevant. These might be important if, for example, the research aims to understand patterns of communication or group dynamics, however our aim was to look at the cognitive processes used when discussing the clinical case. Consequently irrelevant material, that being material that did not relate with the case or the learning taking place which would add unnecessary noise to the analysis, were eliminated. These were mainly detected at the beginning, often before the facilitators were in the room, and end of the sessions. Expressions such as, for example, *"ok let's start", "So are you guys ready?", "ok to the case now" and/or "ok it is all for today",* were common indicators present in change of focus moments either indicating the beginning and end of case discussions. The original transcripts and two versions of the reviewed transcripts were kept, one with 'track-changes' (Microsoft Word functionality) and one only with the relevant material.

Table 4. 11: Total number of words in transcripts before and after removing facilitators' contributions.

Total number of words (with facilitator)	388639
Facilitators contribution	59334
Final transcripts (total)	329305

Pre-software annotation

In our study, we did not wish to add any additional information to the text prior to the automatic annotation process (software annotation). However, in other studies this might be necessary. For example if the speakers need to be identified tags can be placed at this stage. Other common tags that can be added are times, gender, or participants' ID. Usually these tags are placed inside square brackets so that the software recognises that these are not to be considering as words in the text, and will not be counted in word lists, or in any other analysis (Taylor & Gibbs, 2010).

Data Analysis

Data analysis was made by uploading each session transcript into Leximancer and Wmatrix2, electronic content analysis and the corpus analysis software respectively.

For corpus analysis each session was analysed separately and them combined into cycles or years when necessary. In ECA the analysed transcripts were merged into the relevant categories (e.g. all 1st sessions) prior to uploading into the software.

Annotation done by the software adds to each word semantic and grammatical values and classifies them in hierarchical multi-tiered structures of language use. This annotation strategy, in contrast to the coding used in content analysis, does not reflect the particular view or focus the researcher is adopting. It is a generic, fine-grained, data characterisation. At this stage possibilities for analysis are immense, therefore it is necessary to apply to the data a framework that reflects the focus of the research and allows the research questions to be answered. To analyse the cognitive activities taking place during the PBL sessions, the text was used focusing was on the meaning of the words in the text and their particular uses (Kravchenko, 2002). A few categories were selected accordingly to the aims of the analysis: technical language, questions, reasoning and explanations (see sections below).

Analysis of the PBL corpora was performed making use of the classic corpus analysis methods such as frequency lists, keyword lists, semantic and syntactic tags (using the criteria presented above) and concordances. We adopted a type III approach to data analysis as described by Rayson (2008) combining different levels of macro and micro views of the data set. Although an iterative process, we started from a macro perspective, using word frequencies, followed by more in-depth analysis at the concordance level of the key words and categories in our study. This phase was very useful, allowing detection and correction of potential errors of the tagging systems and to gain better understanding of the context in which words, categories, multi-expressions were occurring.

Microsoft Office Excel and SPSS 19.0 were used for further analysis and treatment of results. Statistical analysis was conducted first by using the Log likelihood test available within the corpus analysis software. However, this is a pairwise measure, not allowing a comparison between more than two transcripts simultaneously. Therefore nonparametric statistical tests were used to inspect further relationships between variables under analysis. Nonparametric tests were chosen due to the

relatively small samples used in the study, only two groups, and because in the present study the parameters of the variables in the population were not known.

Technical language

Technical language was identified by using a semantic category part of the UCREL tagging system designated as Body and the Individual and the new dictionary developed. The body and individual category contains words related with healthcare that are frequently used in English conversations (e.g. Pain, hospital, doctor, heart). The Body and Individual subcategories are: Anatomy and physiology; Health and disease; Medicines and medical treatment; Cleaning and personal care and Clothes and personal belongings. The last subcategory was ignored in the analysis performed, as it did not seem of relevance for the topic.

This new dictionary had to be developed as it was noticed many medical terms were not being correctly disambiguated by Wmatrix 2 and were placed in a 'grammatical bin', rather than being allocated to a medical terminology category. The new dictionary added to the USAS tagging system encompassed a list of 43053 words and 1365 multi-word expressions that are used by medical and healthcare professionals (e.g. Ketones, campylo bacterium, intra-orbital) and reflect their knowledge. Figure 4.8 shows the relative frequency of the different categories of words within the new developed dictionary. Detailed description of the development of this dictionary is provided in Appendix 1. The new dictionary encompasses some of the most common informal jargon terms; however, it is not feasible to expect it to cover all the possible acronyms or abbreviated terms students might think of using during their sessions and which get transcribed. Table 4. 12 presents examples of words included in each of the categories. Table 4. 13 presents the results of a comparison of the number of words wrongly identified as grammatical errors (Z99) before and after the new dictionary was added to Wmatrix2. These results show that the development of the new dictionary did in fact improve the accuracy of Wmatrix2 classification of medical terminology.

Figure 4. 8: Distribution of new dictionary words by main category

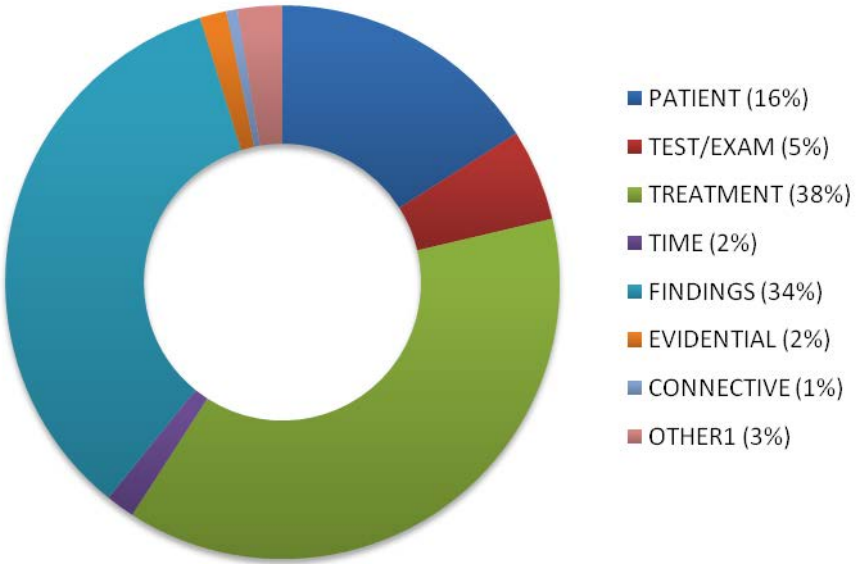


Table 4. 12: Example of 'Common' Medical Words

Example of medical (technical) words added in the new dictionary	Example of medical words in existing software dictionaries
Sputum	breathing
Alveolitis	anatomy
fibrosis	respiratory
oesophagus	lung
larynx	genetic

Table 4. 13: Existing dictionaries vs. New dictionary results

		Existing dictionaries			New Dictionary		
		No. words in Z99	No. medical terms	% of medical terms in Z99	No. words in Z99	No. medical terms•	% of non-recognised medical terms
Session	1	63	30	47.62%	34	5 (2)	14.71%
	2	107	41	38.32%	66	1 (0)	1.52%
	3	51	11	21.57%	38	2 (0)	5.26%

*Terms were misspelled (6) or acronyms (2). The number in brackets represents the absolute number of correctly spelled medical terms although acronyms are tagged as unknown.

Questions

Questions were identified by an analysis of question markers (?) and key words commonly associated with questions was performed (why, how, what, where, when, who). As with technical terms, after frequency identification, a concordance allowed the researcher to review the automatic classification and correct it if necessary.

Reasoning and explanations

This was considered as a single comparison category justified by the results of our validation study. These indicated that the lexical proximity between the verbal expressions of these two cognitive activities could make differentiation between them difficult and error prone. "A" (General and abstract terms), A2 (Affect, and change) and A7 (probability) were used in the identification of explanations and reasoning. These subcategories encompass words that are used in language to express possibility, necessity, certainty, cause, causal relationships, change, make connection between ideas, and how a certain idea will affect others. The POS category refers to subordinating conjugations (CS), and includes words such as for example, "because", "even though" or "so that". This is the category of words that links a subordinate clause to a main clause, or in other words to link a premise to a conclusion (Table 4.14). The presence and frequencies of words in such categories allows us to identify explanations and reasoning, which them the researcher verified by inspection of the text concordances.

Table 4. 14: Example of words and/or multi-word expressions the A2, A7 and CS categories

Example of words in A7 category:	Example of words in A2 category:	Example of words* in the CS categories:
Can/could	to_do_with	Whether
Perhaps	Depends	if_so
make_sure	Happen	Because
Might	Affect/Effect	Although
Sure	Causes	even_if
Probably/ possibility /potentially	something_to_do_with	as_far_as
Clarification	Get/Gets	even_though

*This is a list of a few of the most frequent words in these categories the BNC corpus contains the full list of A2 and A7 and CS contained in the British National Corpora contains several thousand of words and multi-word expressions(PIAO *et al.*, 2005) .

Finally **Electronic content analysis** was used to identify the content of the PBL discussion. This is a necessary step, as the active and self-direct learning nature of PBL assumes there are several different paths that individuals can follow leading to the achievement of these outcomes. Although these learning outcomes were known by the researcher for each of the "observed" content of the sessions provides a much more accurate picture of which content and themes are actually being discussed.

In order to perform electronic content analysis, the software used was *Leximancer*. This is a textual data mining software that allows the identification and extraction of content themes or topics in the text. 'Themes' identified by Leximancer are automatically labelled either based on the most frequent concepts or thesaurus identified in each theme. This system is flexible enough to allow for errors or miss-classifications to be manually corrected case by case. We will follow a similar approach to the one followed recently by Hewett and colleagues (Hewett *et al*., 2009). In this study, focused on doctor-patient communication in multi-specialist departments, the researchers made use of Leximancer software to identify emerging concepts and 'themes' from interviews with 45 doctors. This analysis was then followed by an interpretative phase based on the authors' extensive readings of the materials. This was also done to less detailed extent in our study, as our aim, is not to characterise complex communication patterns but rather understand what was the explicit content of the discussion in the PBL sessions. Additionally, as this was not the main methodology used in the study and considering the scope and space limitations of the present thesis, it was decided to focus presentation of results on the yielded 'themes' with the inclusion of some extracts from the transcripts as examples. However, during the data analysis phase the transcripts were inspected using the software's in-built tool that allows seeing the words contained in each 'theme' in their context. This was done to further understand the 'theme' classification and account for possible errors of classification due to context ambiguity. Leximancer software is very flexible allowing the researcher to determine what type of queries to be performed, but also what is the format of results and how those are to be exported. One of these customised features is the number of visible topics, or 'themes' and their granularity generated. Granularity in electronic content analysis means the number of concepts included within each topic or 'theme', therefore the higher the number of visible 'themes' the lower the number of concepts nested in those 'themes', that is their granularity. Granularity in Leximancer varies between

100/100, where all concepts would be merged into a single 'theme' and 1/100 where one 'theme' per concept identified in the text would be generated.

C.8 Criteria For Ensuring Rigour And Quality

As described above, the current study used electronic text analysis methodologies, corpus analysis and electronic content analysis to analyse recordings of discussions taken place during PBL sessions. We adopted a mixed-methods methodology (Creswell & Clark, 2011), making use of a type III analysis of the corpora to search for answers to the research questions (Rayson, 2002). Therefore, the criteria to ensure quality and rigour of the study must reflect theses methodological choices.

Several have suggested questions or frameworks to identify quality and rigour criteria in mixed-methods studies (Leech & Onwuegbuzie, 2007; Heyvaert *et al*., 2011; Creswell & Clark, 2007, 2011; Migiro & Magangi 2011). The aspects identified by these authors can be summarised as:

Assure coherent choice of methods (qualitative and quantitative) according to the research questions;

To describe and discuss possibility of conflicts and tensions between the two qualitative and qualitative parts of the research

To ensure that qualitative and quantitative parts of the study respect criteria for judging quality within each of the reported paradigms.

Our adopted criteria also reflected the nature of data (corpus) involved in the study, by ensuring that both collection and treatment of corpora was complying with necessary quality requirements of collection and analysis of corpus in linguistics, mainly the representativeness, size, and comparability of corpora. The corpora representativeness was ensured by recording PBL discussion covering a range of different PBL cases, within two different modules per group and sessions involving different facilitators (one clinical and one non-clinical per group). As our aim was to build corpora of spoken PBL discussions no additional materials from other than spoken sources were included in the corpus.

Corpus size is related with the research aims and purposes and to which the information in the corpora is large enough to be able to draw appropriate conclusions. As our aim was not to be able to make generalisations across all PBL curricula, but rather use the results of the analysis of these two groups to inform the

discussion of possible differences found in a comparative study (study 2) conducted with the remit of the current research, the size of the corpora collected was considered sufficient for that purposes. Comparability of the corpus is extremely important, especially in the data analysis phase. A couple of considerations were accounted for here, first we made sure the same type of data (discussion recordings) was collected from both PBL corpus (year 1 and year 2), so those could be compared. In both, data collection followed exactly the same procedure, also in the pre-analysis phase the aspects removed (facilitators contributions and 'irrelevant material') from the corpora were the same for both groups, additional validation and correction procedures carried out by the researcher were exactly the same for both cases. Finally in the analysis phase it was ensured that corpora comparison was always done against spoken corpora (BNC-Spoken speech) and not corpora involving other types of textual materials.

Finally ethical aspects and considerations, described in the beginning of this section were also a fundamental element of quality and rigour assurance in the study. Confidentiality of data was strictly kept along the study. Participants were not identified in the transcripts and any details or any other personal information revealed during the PBL discussion was removed from the transcripts by manual inspection during the 'cleaning' data preparation phase. Video data and audio was stored in a password-protected external hard-drive in a locked cabinet to which only the researcher and the supervisor of this research would be able to access.

Chapter 5: Analysis of PBL Sessions (Study 1) - Results

"Language is a window to the human mind"

(Pinker, 2007)

Chapter summary

The last chapter presented a validation of a methodology (corpus analysis) used in order to analyse cognitive activities taking place during PBL session, how it was combined with an electronic content analysis in order to gain further understanding of such processes and a description of the steps taken in order to answer the research questions posed by the current work. In that chapter we described the philosophical approach underpinning the choice of such methodologies and purpose of their use in the current research study. Research questions, sampling, data collection and data analysis procedures were also presented.

The present chapter presents the results, discussion and limitations of the application of the aforementioned methodology.

First the general descriptive data regarding all the data collected and analysed will be presented. After the topics of discussion (or 'themes') identified through electronic content analysis will be presented, to answer the question "what was being discussed in the PBL sessions?" Then results regarding the corpus analysis of technical language, analysis of questions, reasoning and explanations will be presented. These results will answer the questions about "how were the students approaching the cases? What were the frequencies of questioning, technical words, and reasoning and explanations in these sessions?"

Following we will discuss the results of both studies, focusing on what can they 'tell' about how early stage medical students approach clinical problems. Study limitations will also be presented and discussed, and main

A. Results

A.1 The Structure Of The PBL Sessions

In the Graduate Entry Medical School in the University of Nottingham the PBL sessions are structured so that the first session should be aimed at introducing the scenario often with videos of patient presentations and a follow up discussion with the facilitators, where students can be provided with additional information on the patient history. In this session students also generate learning objectives and explore issues to be discussed further. The second session is aimed at providing accounts and explanations of relevant learned material. In this session the facilitator can provide the students with additional data from clinical investigations to be explored. And a final, third session dedicated to synthesise the case and generate conclusions, discuss and agree diagnosis and treatment plans.

A.2 Sample Characterisation

Our sample was constituted by two PBL, each contain seven students and a facilitator present at the time of each session. The gender distribution was similar in both groups, with a higher percentage of male (57%) students. This is contrary to the UK national picture, where by a higher percentage of females (57%) are admitted to graduate Entry Medical Schools. Demographics and application information presented in Table 5. 2, show a high degree of similarity between first and second-year groups. The average age at entry was 27 and 28 for the first and the second-year, respectively, with the most frequent age group age (for both) being 20 to 25 years old, and only one person reporting an age higher then 35 years old. Also regarding previous degrees the groups can also be considered similar, although the students have different degrees, generically the distribution across non-science, science (non-biomedical) and biomedical degrees is similar as shown in Table 5. 2. Some differences were noticeable in the distribution of GAMSAT scores; however, the difference between the averages is very small.

Table 5. 1: Gender distribution by PBL groups

	Female	Male
First-year Group	43% (3/7)	57% (4/7)
Second-year Group	43% (3/7)	57% (4/7)
UK Graduate Entry Medical schools	57%*	43%*

*National reference values from 2003-9 as published by (Garrud 2011)

Table 5. 2: Age, GAMSAT and Previous Degree distribution by PBL group

Demographic and pre-entry data	Year 1	Year 2
Age		
[20-25[	57% (4/7)	43% (3/7)
[25-30[	29% (2/7)	29% (2/7)
[30-35[	0	14% (1/7)
[≤ 35]	14% (1/7)	14% (1/7)
Average age	27	28
Degrees		
Biomedical	43% (3/7)	43% (3/7)
Science (other non biomedical)	43% (3/7)	43% (3/7)
Non Science	14% (1/7)	14% (1/7)
GAMSAT score		
[60-62[	14% (1/7)	14% (1/7)
[62-64[	57% (4/7)	29% (2/7)
[64-66[	0	29% (2/7)
[66-68[	29% (2/7)	14% (1/7)
[≤ 68]	0	14% (1/7)
Average GAMSAT Score	63	65
Total number students per group	7	7

* This table is based on information on individual participants' provided in Appendix 2: Table AP2 2.

A.3 Descriptive Results

The present study collected more then 60 hours of video and audio recordings of PBL sessions. A total of 41 sessions were recorded, adding up to a total of 388639 words as shown in table below. A larger number of first-year sessions were recorded due to the process of recruiting volunteers for the study. The announcement to the second-year students was delayed by external factors, leading to a later start of the recording process. Nevertheless more than 20 hours of PBL sessions were recorded.

Table 5. 3: Data collection general information (full information on Appendix 2: Table AP2.1)

Year (Groups)	First session recorded (date)	Last Session recorded (date)	N of sessions/hours* per week	N of modules recorded	N of sessions
Year 1	14/11/2008	03/02/2009	3/ 4.5h	2	26
Year 2	02/12/2008	23/01/2009	3/ 4.5h	2	15

*According to 2008/2009 timetable **Full data per session is presented in Appendix 2: Table AP2.1.

Of interest is the higher density of words per hour in first-year students. It is important to notice here that these are values based on previously cleaned transcripts from which non-relevant material to the case discussion and facilitators contributions was removed. This was confirmed by a visual inspection of the video recordings where it was observed that the first-year group was more "talkative" then the second-year group.

Table 5.4: Total number of sessions, hours of recording and number of words analysed (full information on Appendix 2: Table AP2.1)

	Total number of sessions recorded	Total duration of the sessions according to the timetable in hours	Total number of hours recorded	Total number of words*	Average Word density (w/h words by hour)**
Year 1	26	39.5	39.1	239119	6115 w/h
Year 2	15	24.5	21.7	90186	4156 w/h
Total	41	64	60.8	329305	5416 w/h

*Cleaned transcripts ready for analysis **Word density is calculated by dividing the total number of words by hours of recording in each session

The GEM programme, as described in the chapter dedicated to the comparative study of the curriculum, is organised by modules that cover areas of study for the 18 months of the programme. Each module is based on a body system, and contains a variable number of clinical scenarios (cases) that the students explore in the PBL sessions, with additional lectures and clinical skills sessions.

Table 5.4 shows the distribution of the number of words captured per module and clinical cases/scenario.

Each case follows the steps of the clinical cycle starting with a first session dedicated to the patient presentation and history and an exploration of questions to ask to the patient. A second session is dedicated to an interpretation and analysis of patient physical examination, plus a discussion of clinical investigations to perform. A final session is usually focused on summarising the findings, identifying a diagnosis or a

differential diagnosis; defining a management plan, that can or cannot include medications and prescriptions, and a summary of the learning outcomes achieved. Each case, although having a previously defined diagnosis, is designed to allow the students to explore, during discussion, alternative differential diagnoses and related other conditions. The diagnosis summary on Table 5.5 should be interpreted as an indicator of the area of diagnosis that is being discussed. For example the case '*To cap it all*' it would expect the students to be discussing and exploring not only *Haemophilus influenza* but other types of community acquired influenza and other possible related differentials.

The four last second-year scenarios are described as *Integrative,* these scenarios cover several of the body systems explored by the students during the previous (approximately) 14 months of GEM programme. These are complex problems involving infections, syndromes and other pathologies with the aim of making the students mobilise previously acquired knowledge and skills in order to be able to explore and reach conclusions about the case. These cases are the last cases of the PBL programme, following these the students have their assessments and if they pass they start their clinical practice phase 1.

The information shown in the above table is based on material provided by the facilitators' guides which contain recommended details of themes and objectives to be discussed in the sessions Due to the nature of the PBL scenarios, students' learning processes are flexible and (self) directed by the group, hence not possible to predict exactly. For this reason, in this study an electronic content analysis was carried out, and its results are described in the next section.

Table 5.5: Modules, case title and diagnosis of recorded sessions

Year	Module	Case (cycle)	Diagnosis summary	Total hours recorded *	Total number of words**
1	Respiratory	To cap it all	Haemophilus Influenza	3.7	13781
1		It's affecting my football	Asthma	4.1	32070
1		Can we fix it?	COPD	4.6	35575
1		Out of Africa	CA lung	4.2	23323
1	Cardiovascular	Life is unpredictable	Associated ventricular septal defect	4.4	28282
1		Be still my beating heart	Atrial fibrillation with underlining heart disease	4.2	26560
1		It can wait until Monday	Non ST elevated myocardial infarction (NSTEMI)	4.8	31639
1		Old and crusty	Symptomatic aortic stenosis	5.2	24291
1		Benign or not benign?	Widespread vascular disease in elderly	3.9	23598
2	Neurosciences	Only the lonely	Trigeminal neuralgia.	4.5	20752
2		Twisted Sister	Ischaemic stroke, atherosclerosis	3.4	19633
2		A Shaky Serve	Ipsilateral loss of motor function	2.3	1714Δ
2	Integrative	Hard labour	non-massive pulmonary embolism	3.1	15600
2		Blood Brothers	Chronic myeloid leukaemia (CML)	1.7	7051Δ
2		More than just a rash	HIV infection	4.7	14149
2		Put your foot down	'The Diabetic Foot' syndrome	3	14380

Δ Only one session of this module was recorded *Full information per session presented in Appendix 2: Table AP2.1 **After cleaning processes

A.4 Electronic Content of The Sessions - What Were The Students Talking About?

This section presents the results of the electronic content analysis conducted on the transcripts of the PBL sessions recorded. This analysis allows an identification of the main topics or 'themes' being discussed in the PBL sessions, which can be compared to the information provided by the facilitators' guidelines and the case summary for each of the sessions. The guides are not accessible to the students, only the facilitators have previous access to this information. The students only receive information regarding the patient's initial presentation and further data when requested by the group. Learning objectives are provided at the end of the three sessions to allow them to revise and relate to further study of any areas that were not covered. This analysis is complementary to the main data analysis conducted thought corpus analysis, and allows for a better interpretation and discussion of the latter.

Table 5.6 and Table 5.7 compare the main themes identified by our analysis with the diagnostic summary presented in the facilitators' guides for the sessions, for the first and second-year respectively. Overall the themes discussed correspond to the areas defined by the modules, being related to body systems, organs, signs, symptoms, possible diagnosis and possible consequences of diseases. These were focused on between one and three main themes encompassing the majority of the discussions taking place during each session when a granularity[34] of 50/100 is selected. Although during the analysis all the granularity levels were used to gain understanding of the data, the 50/100 granularity level proved to be a suitable choice considering the balance between specificity, differentiation and meaning of the 'themes'. This choice means that 'themes' grouped more then one identified concept, there is a sufficient number of different 'themes' and each of those are coherent units that can be meaningfully analysed, which would not be the case if all the concepts in the text were to be grouped in one large 'theme'.

34. Granularity in Leximancer defines the number of concepts clustered within a 'theme'. It varies between 100/100, where all concepts would be merged into a single 'theme' and 1/100 where one 'theme' per concept identified in the text would be generated. More details in Chapter 4 Section: Data Analysis.

Table 5.6: First-year session's main themes expected vs. observed

Module	Case (cycle)	Diagnosis summary (Excepted theme)	Case summary (Excepted theme)	Main themes identified in the sessions* (observed themes)
Respiratory	2.2.To cap it all	Haemophilus Influenza	"This case concerns a lady in a nursing home who has a community-acquired infection (Haemophilus Influenza)" (Ongoma, 2008, p. 1)	Disability/Incapacity, Personal Relationships **[2] Lung, Sputum (equivalent) [3]
Respiratory	2.3.It's affecting my football	Asthma	"This is a case of a young boy who suffers from asthma. His condition is brought on by exposure to allergens and exercise and is accompanied by atopic eczema" (Hughes *et al.*, 2008, p. 1)	Breathing (main) question and time [1] Expiratory (main) Months [2] Capacity (breathing capacity) [3]
Respiratory	2.4.Can we fix it?	COPD	"This is a case of COPD exacerbation in a middle-aged male smoker who has come to the doctors because of increasing difficulty breathing and fatigue" (Hughes, 2008a, p. 1)	Bronchitis (main), respiratory [1] Obstructive (main) night [2] Breath (main), depends, year [3]
Respiratory	2.5.Out of Africa	CA lung	"This is a case of an elderly man presenting with signs and symptoms of chest disease which could be infective or neoplastic. This leads the students to explore both carcinoma (CA) and Tuberculosis (TB) of the lung" (Seigel, 2008, p. 1)	Diseases (main), tiered [1] Symptoms** (main), die, feel [2] Cancer (main), Case [3]
Cardiovasc	3.1. Life is	Ventricular	"This is a case of a baby who is born with Down Syndrome and	Down syndrome (main) [1]

ular	unpredictable	septal defect in baby with down syndrome	is found to have a small sized ventricular septal defect" (Manning, 2008a, p. 1).	Down syndrome, VSD, ECG [2] Learning experience** [3]
Cardiovascular	3.2. Be still my beating heart	Atrial fibrillation with underlining heart disease	"This is a case of a patient who experiences palpitations. The case also raises issues of cultural diversity as the patient is a practising Sikh" (Manning, 2008b, p. 1).	Palpitations [1] ECG [2] ECG (main), [3]
Cardiovascular	3.3. It can wait until Monday	Non ST elevated myocardial infarction (NSTEMI)	"Pathophysiology and natural history of arterial disease with emphasis on coronary artery disease: Reduction of cardiovascular risk factors; Electrocardiography normal and ischaemic; Hyperlipidaemia" (Waterhouse, 2009a, p. 1) .	Coronary [1] Angina (main) [2] Disease (main) [3]
Cardiovascular	3.4. Old and crusty	Symptomatic aortic stenosis	"Pathophysiology of the failing heart and clinical and laboratory assessment of heart failure."(Waterhouse, 2009b, p. 1)	Oedema [1] Heart [2] Heart failure [3]
Cardiovascular	3.5. Benign or not benign?	Widespread vascular disease in elderly	"This is a case of a patient with widespread vascular disease associated with long standing hypertension, poorly controlled blood pressure and a plethora of cardiovascular risk factors." (Manning, 2009, p. 1)	Vein (main), people [1] Pain (main), cholesterol ** [2] Anatomy, smoking habits ** [3]

* Each line corresponds to a session in the case. Number in square brackets indicate the session number []** 'Themes' manually corrected by the researcher based on the analysis of the themes node and meanings. Due to need for disambiguation or poor automatic labelling of themes, not corresponding to the best term to describe nodes grouped.(main) - Identifies main theme in the session, being related with more then 70% of the content of the session. [Number]- Identifies the PBL session from the case.

Table 5.7: Second-year sessions main themes expected vs. observed

Module	Case (cycle)	Diagnosis summary (Excepted theme)	Case summary (Excepted theme)	Main themes identified in the sessions* (observed themes)
Neurosciences	8.4. Only the lonely	Ischaemic stroke, atherosclerosis	"This is a case in which an elderly lady has had an ischaemic stroke involving one of the branches of the middle cerebral artery. The presentation is typical, involving areas of the brain supplied by that artery."(Rowan-Robinson, 2008, p. 1)	Channels (main), Because [1] Suicide, muscles (equivalent) [2] Trigeminal, nerves (neuralgia) [3]
Neurosciences	8.5. Twisted Sister	Ischaemic stroke (middle cerebral artery)	"This is a case in which an elderly lady has had an ischaemic stroke involving one of the branches of the middle cerebral artery. The presentation is typical, involving areas of the brain supplied by that artery.'(Mclaughlin, 2008, p. 1)	Motor (main), stroke [2] Nerve (main), protein [3]
Neurosciences	8.6. A Shaky Serve	Ipsilateral loss of motor function	"This case centres around a 55 year old man who has the progressive movement disorder Parkinson's disease." (Hughes, 2008b, p. 1)	Antagonist (main) tremor [2]
Integrative	9.1. Hard labour	non-massive pulmonary embolism	"This case acts as a prompt to revisit and to integrate prior learning about pregnancy, childbirth and embryological development."(Garrud, 2009, p. 1)	Acid (main), liver [1] Respiratory (main), factor [2] Impulse, pressure [3]

Integrative	9.2. Blood Brothers	chronic myeloid leukaemia (CML)	“This case deals with the haematological and immunological issues around leukaemia. It will provide the opportunity for students to explore blood cells and their progenitors; anaemia; leukaemia; immunity and immunosuppression; issues of transplantation.” (Hagan, 2009, p. 1)	Cells (main), B12 [2]
Integrative	9.3. More than just a rash	HIV infection.	“There are several layers of diagnosis here. The patient has herpes zoster with evidence of dissemination, oral candidiasis, and lower respiratory tract infection. In a 28 year-old, previously apparently fit patient, these diagnoses are insufficient to explain the symptoms.” (Irving, 2009, p. 1)	HIV (main), deficiency [1] DNA (main), cells [2] Respiratory (main), gram [3]
Integrative	9.4 Put your foot down	'The Diabetic Foot' syndrome	“This case presents students with opportunities to review their knowledge of Type 2 Diabetes and complications associated with diabetes.“ (Jennings & Dhindsa, 2009, p. 1)	Blood, insulin [1] Question (main), pressure [2,3]

* Each line corresponds to a session in the case. Number in square brackets indicate the session number []** ‘Themes’ manually corrected by the researcher based on the analysis of the themes node and meanings. Due to need for disambiguation or poor automatic labelling of themes, not corresponding to the best term to describe nodes grouped.. (main) - Identifies main theme in the session, being related with more then 70% of the content of the session. [Number]- Identifies the PBL session from the case.

Discussions are not centred on a particular diagnosis but rather on the underlining causes and presentations. The themes identified in a first-year session show that students' discussions tend to focus more on exploration of symptoms, signs, clinical investigations and an understanding of underlining causes. Cases 2.4, 2.5 and 3.3, are the only cases where a particular diagnosis was identified as being central to the discussion. However, only in one of these did the actual diagnosis of the case emerge as a theme, namely the last respiratory case: Out of Africa (2.5). This case diagnosis was lung cancer and cancer was one of the main themes of the third session, in the other two the pathologies being explored were possible differentials diagnosis for the case. A similar picture emerged from second-year data, in only one case. More than just a rash (9.3), the case diagnosis emerged as one of the themes being discussed. In this case, the discussions focused more on the mechanisms and underlining process of disease than on the actual diagnosis or prognosis. This observation seems to be aligned with a core assumption of the PBL curricula, that the cases aim to promote learning in a broader module-wide area.

The case 8.6, Shaky Nerve, was the only case where treatment options and drugs/prescriptions were identified as being central to the discussion ("Antagonists" theme). In case 2.2 psychosocial aspects were being discussed by the students, mainly related with the impact of the symptoms in the patients' lives, grouped under 'disability/incapacity' theme and the personal relationships of this elderly lady (patient). Nevertheless, this is the only theme concerning these aspects that emerged in our analysis, in both first and second-year discussions. The sessions map onto the expected PBL structure with first-year groups but not second-year groups. At GEM the PBL sessions have a previously defined structure, allowing for the group to have some autonomy and flexibility in managing their discussion time. A first session should be dedicated to history, a second session should be focused on additional questions and clinical data, and a third and final session is usually concerned with the final diagnosis and possible treatment plan. Overall the themes identified in first-year data relate closely with this structure, but not the ones identified in the second-year.

First-year data show that the discussion in the first session tends to encompass elements of case presentation. Case 2.2 is an example with the name of the patient being one of the identified themes. Case 3.1, Life is unpredictable, is another example, where the theme "down syndrome", identified as a main theme of the discussion is part of the initial presentation of the case. Within this first-year group it

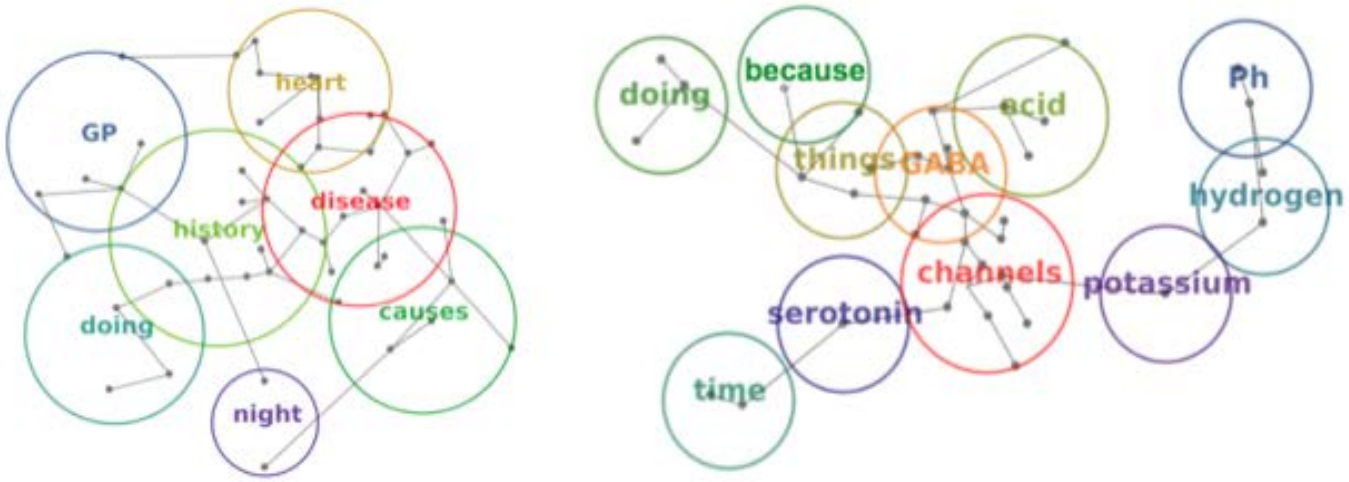
is also possible to see that the second sessions are, mainly, dedicated to discussions related to clinical examination or data derived form observations. The third session however seems to be a continuation of the second session, with the exception of the case 2.5, Out of Africa, where the diagnosis is the main theme of discussion. Case 3.1, Life is unpredictable, also shows an exception, in the third session of this case students are discussing their learning process, as shown by the themes 'session' and 'case'.

Second-year themes do not do seem to establish such close connections with the GEM recommended session structure[35]. The main themes emerging from this dataset show a focus on the understanding of mechanisms, processes and structural or biological functions, that relate to the case. Indeed the increased focus on the basic science mechanisms and pathophysiological explanations of the underlining mechanisms of disease and body systems was also supported by the manual inspection of the transcripts. However, this does not mean, *per se*, that the defined PBL structure has not been adopted at all. Rather it is not possible to identify it in the present analysis. A tentative reason maybe that the group is adopting a more flexible structure in order to focus on their learning needs rather then simply following the typical PBL case structure of case presentation-session 1, data discussion on session 2 and diagnosis and treatment on session 3.

Figure 5.1 to Figure 5. 3 show a visual representation of the main themes and the connectivity of the clustered concepts (grey circles and lines) per session one, two and three in both years of study yielded by Leximancer.

[35] Details of the PBL structure adopted by the school were presented on Chapter 3.

Figure 5. 1: Main themes and connections in the first PBL session: first-year group vs. second-year group



First-year sessions

Second-year session

Note: 'themes' presented in this figure were generated from a merge transcript of all PBL sessions in analysis (all 1st sessions for year 1 and all 1st sessions for year 2). This procedure allows the software to identify the most frequent themes across all the sessions, rather then summing up frequent themes per sessions, causing errors in data due to equal corpus size. Connection lines represent co-occurrence of the concepts in the text. Grey dots represent frequent concepts that are aggregated within the main identified 'themes' inn the centre of the circle. Grey lines provide a visual representation of co-occurrence of such concepts. These two elements allow identification of relationships between 'themes' in the map, by looking at how frequently the concepts that form a 'theme' co-occur with concepts from other 'themes'. Overlapping between the circles represents this relationship between 'themes'. The 'themes' here identified are not a sum of the themes identified in each session. 'Themes' identified in the red circles cluster the higher number of concepts in the data. Position in the map revels the association between the theme and other themes; the more central a theme is positioned the higher the number of associations with the other identified theme. Association between themes is calculated based on the co-occurrence of concepts clustered within different themes. Definition and examples of utterances in the themes are provided in Appendix 2: Table App 2.3.

Leximancer analysis identified the dominant themes within year one students 1st PBL sessions as: Disease, history, causes, heart, GP doing and night, with history being the most connected theme and disease the most frequent. The 'themes' identified in the second-year sessions were: Channels, potassium, hydrogen, pH, GABA, Acid, Serotonin, time, things, because and Doing with GABA being the most connected 'theme' and 'channels' the most frequent (Figure 5. 1).

For first-year students the identified themes seem to reflect the aims defined by the school for these sessions, which suggest these sessions (1st) are focused on history taking, exploration/brainstorm of possible processes and causes of the patient's signs and symptoms. Additionally, the analysis revealed a presence of concepts related with cardiovascular system in the discussions ('heart'), which is aligned with the fact that these sessions were part of the cardiovascular and respiratory modules.

The 'theme' history is central to the first sessions in the first-year groups, with the concepts in this theme co-occurring frequently with concepts in other themes identified in the data set. Centrality and high connectivity of this 'theme' demonstrate the high importance of history taking and case initial information for the discussions taking place during these first sessions for year one students. The theme labelled 'disease' included words such as illness, infection, disorder, defect, and influenza, among other disease related concepts. The high frequency of this theme indicates that a high percentage of the discussion is focused on diseases and diseases processes. Its proximity to the history 'theme' indicates that these discussions involve information from the patient history. Equally the proximity with 'heart' 'theme' and 'causes' indicates that the students' discussion of diseases frequently involves aspects related with cardiovascular system ('heart' theme) and exploration of other possible causes for diseases ('cause'). Finally, the other 'themes' identified 'GP', 'night' and 'doing' seem to be less central or frequent in the discussions, still encompassing concepts that co-occurred frequently with those within the history 'theme'.

The second-year discussions are more complex, involving a higher number of themes and are focused on specific scientific concepts such as channels, potassium, GABA, however with less interconnected themes and not a single central theme. In these, history does not emerge as a 'theme'. The three main groups of 'themes' identified reflect discussions focused on understanding the pathophysiological processes of the clinical problems presented by the cases. Some of the 'themes' identified correspond

to scientific terminology, such as 'channels' or 'pH' and others, as 'because'. For example a cluster potassium, hydrogen and pH, clearly demonstrates that the students are focusing on biochemical and physiological processes such as acid-base balance and acid-base disturbances. Of notice is also the theme labelled 'things'. This 'theme' denotes the colloquialism of the conversation, but also reflects the frequent use of the word thing and its variations to replace technical terminology or to replace explanations of processes, revealing that students' technical lexicon is still being formed as exemplified below.

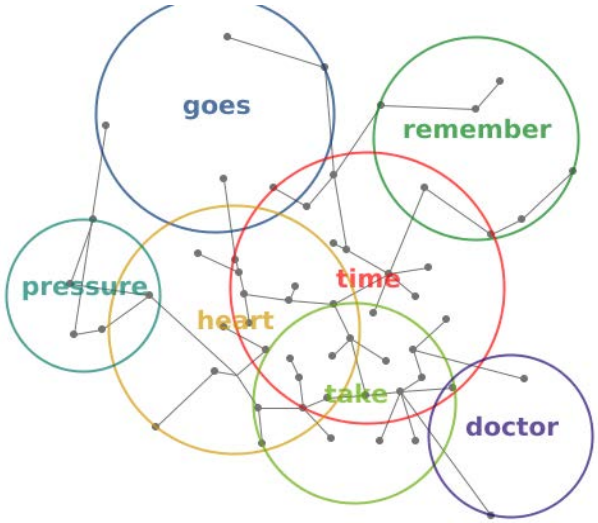
"I thought most uptake channels tended to be location specific but not substrate specific, so you could have, like, serotonin reuptake channels but they can reuptake other things as well to a lesser degree." [02.8.4v1]

"Reboxetine causes dizziness... they seem okay apart from tricyclic anti-depressants are the same things, noradrenaline is…" [02.8.4v1]

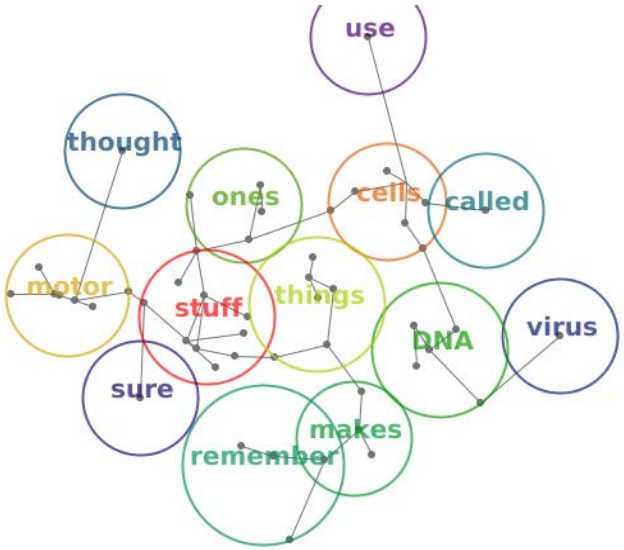
Two main facts can possibly contribute to explain these differences between the first and second-year sessions. On the one hand, the history-taking process should be covered in more detail on the first-year of the GEM programme. It is during this time that the students learn how to make relevant enquiries and approaches to patients and their families in order to collect information. They learn how to distinguish between relevant and non-relevant questions among other aspects of the history taking process (University of Nottingham 2011a). Therefore it is more relevant in PBL discussions.

On the other hand, second-year students, having completed 12 months of PBL sessions, are much more familiar with the process. This may make the group become more responsible for organising the structure of discussion so as to cover the areas they need. As they are about to enter clinical practice it is expected that the students will make an additional effort to understand some crucial medical concepts and relationships between them. This will lead to the sessions covering a higher concentration of scientific concepts as themes, because the students are talking more about them, and not being so case focused as the first-years.

Figure 5. 2: Main themes and connections in second PBL sessions first-year group vs. second-year group



First-year sessions



Second-year session

Note: 'themes' presented in this figure were generated from a merge transcript of all PBL sessions in analysis (all 2nd sessions for year 1 and for year 2). This procedure allows the software to identify the most frequent themes across all the sessions, rather then summing up frequent themes per sessions, causing errors in data due to equal corpus size. Connection lines represent co-occurrence of the concepts in the text. Grey dots represent frequent concepts that are aggregated within the main identified 'themes' inn the centre of the circle. Grey lines provide a visual representation of co-occurrence of such concepts. These two elements allow the identification of relationships between 'themes' in the map, by looking at how frequently the concepts that form a 'theme' co-occur with concepts from other 'themes'. Overlapping between the circles represents this relationship between 'themes'. The 'themes' here identified are not a sum of the themes identified in each session. 'Themes' identified in the red circles cluster the higher number of concepts in the data. Position in the map revels the association between the theme and other themes; the more central a theme is positioned the higher the number of associations with the other identified theme. Association between themes is calculated based on the co-occurrence of concepts clustered within different themes. Definition and examples of utterances in the themes are provided in Appendix 2: Table App 2.3.

A similar pattern can be identified on the analysis of the second sessions, with second-year discussions being more complex and encompassing a higher number of themes. Saturation of scientific concepts is still higher in the second-years than in first-year discussions, although the difference is smaller then in the first sessions.

First-year discussion seems to be largely focused around *time* and *heart. The heart* theme, as before indicates that the cardiovascular system is being explored in the discussions. *Time* encompasses words and expressions related to time mainly discussing physiological processes, as exemplified by the quote below, and a few words/expressions related to the time of clinical investigations/result.

"They live for a very short time but they've got loads of telomerase so that they can, you know, lots of cell division and lots of cell repair." [01.2.2v2]

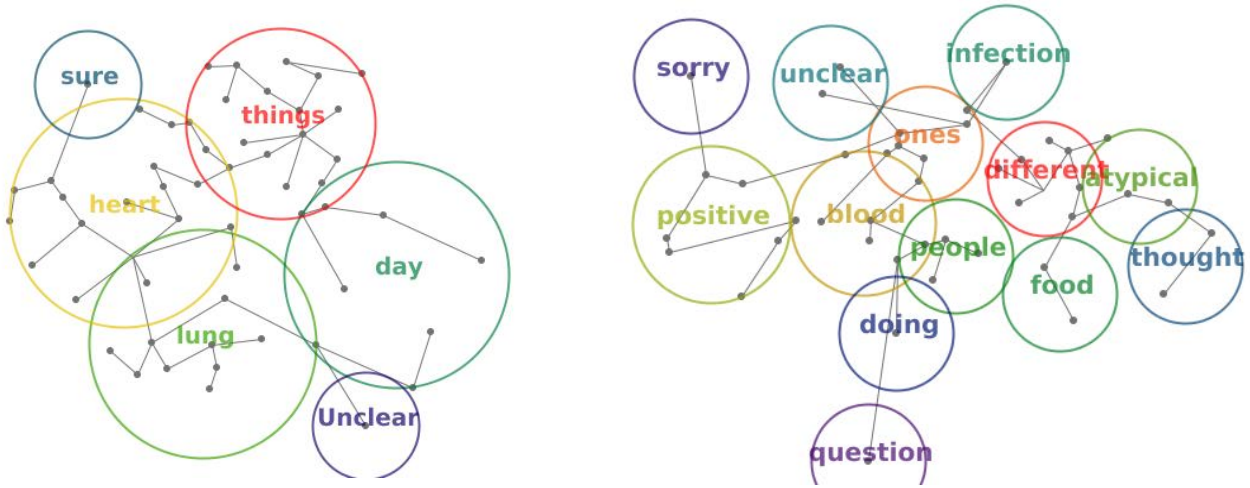
As before, on the first session, the second-year discussions are not as interrelated as the first-years, also due to the higher number of themes emerging. Of notice, here a few generic themes are identified, mainly '*stuff*' and '*things*' themes. This is explained by the frequent use of 'replacement words', words and expressions as shown in the quotes below that are either the words for 'stuff' or 'things' or their synonyms. These words are used in the students' discussions to replace nouns that the students do not seem able to recall at the moment of the discussion.

"And what's this white stuff on the outside?" [Question done while analysing a figure in a text book] [02.9.2v2]

"we're not expecting her know name of genes that predispose it but what kind of evidence has it got that in order to check these things ..." [02.8.4v2]

Additionally in both years "*remember*" has been identified as a theme. By a more detailed analysis of the words in the theme it is possible to conclude that on the one hand both groups of students are making connections between this session and the information, discussion and conclusions they had researched in the previous session of the case. On the other hand, they are identifying aspects considered crucial for their learning objectives and that they should learn/memorise for future use, using expressions such as "*I must remember that for....*".

Figure 5. 3: Main themes and connections in third PBL sessions first-year group vs. second-year group



First-year sessions

Second-year session

Note: 'themes' presented in this figure were generated from a merge transcript of all PBL sessions in analysis (all 3rd sessions for year 1 and year 2). This procedure allows the software to identify the most frequent themes across all the sessions, rather then summing up frequent themes per sessions, causing errors in data due to equal corpus size. Connection lines represent co-occurrence of the concepts in the text. Grey dots represent frequent concepts that are aggregated within the main identified 'themes' inn the centre of the circle. Grey lines provide a visual representation of co-occurrence of such concepts. These two elements allow the identification of relationships between 'themes' in the map, by looking at how frequently the concepts that form a 'theme' co-occur with concepts from other 'themes'. Overlapping between the circles represent this relationship between 'themes'. The 'themes' here identified are not a sum of the themes identified in each session. 'Themes' identified in the red circles cluster the higher number of concepts in the data. Position in the map revels the association between the theme and other themes; the more central a theme is positioned the higher the number of associations with the other identified theme. Association between themes is calculated based on the co-occurrence of concepts clustered within different themes. Definition and examples of utterances in the themes are provided in Appendix 2: Table App 2.3.

Finally, the third sessions present a similar difference in complexity and interconnection between a number of themes observed in the previous sessions (higher for second-year with less interconnectivity). In second-year discussions the saturation of scientific concepts is less than in the previous sessions, being almost the same as in the first-year discussions. However, the centrality of the theme 'blood' indicates that the discussions on this final session are revolving around

Interestingly the theme 'lungs', indicating that the discussion is exploring more aspects related to lung function and the respiratory system, appears in this session as one of the main themes on the first-year group discussion. This is the area of the three-first cases. However, from the analysis of all the sessions it appears to become more common in the third session, while the 'heart' theme is constant across all the three sessions. A possible explanation, based on the manual inspection of the context of the themes, seems to indicate that nature of the case has an impact in the strategies chosen by the students to explore it. In respiratory cases students seem to be discussing lung function and respiratory system more in the third session while in the cardiovascular cases these topics seem to be distributed between all sessions. However, further analysis would have to be performed in order to confirm this possibility.

The above results seem to confirm that overall second-year sessions have a more complex structure with more themes being identified. There is also a noticeable difference in the nature, or type of themes, with the second-year themes having overall more scientific concepts.

The results presented here will be discussed further in the section discussion/conclusions along with the results form the main analysis method: *corpus analysis*. The following sections will present the results of corpus analysis of technical language, questioning and reasoning by session, case (cycle), session type and years. Additionally a final section will describe the results of the exploration of early clinical reasoning strategies used by the medical students. This section will give the measured frequencies of questions, explanations and reasoning, used by the students in comparison to the evidence from the literature on experts and novices when engaging in the process of solving clinical cases.

A.5 Corpus Analysis – How Are The Students Approaching The Cases?

The present section presents the results of the corpus analysis of technical language questions, explanations and reasoning in the recorded PBL sessions.

From a learning perspective, the ability to ask relevant questions, to explain ideas and to make connections between new and acquired concepts are crucial to promote the accommodation, and assimilation of new concepts, leading to significant long-lasting learning (Bloom, 1992). These skills promote integration of concepts into gradually more complex networks contributing to knowledge organisation in memory and to clinical reasoning development (Norman, 2005; Bowen, 2006). It is important, therefore, to analyse how these skills are being used in the PBL sessions as a way of gaining some understanding of what is the early basis of clinical reasoning. We aim to understand how the students are using technical language, questions, explanations and reasoning to discuss the clinical cases and what these elements can tell us about the beginning of the development of clinical reasoning in the PBL context.

Technical Language

In the present section the results of the corpus analysis of technical language will be presented. Analysing students' use of medical terms can help to gain insight into the knowledge encapsulation process, a crucial step in the development of clinical reasoning (Boshuizen et al. 1992; Schmidt & Rikers 2007; Rikers et al. 2000). The knowledge encapsulation process is defined as the process by which students' basic biomedical concepts are integrated into complex knowledge networks. These networks are formed by groups of concepts with close links between each other designated as clusters and by the relationships these clusters establish with other clusters. As the students learn biomedical knowledge these clusters and their relationships become increasingly denser, encompassing a higher number of clusters, a higher number of concepts per cluster and a higher number of possible relationships between them (Boshuizen, 2003).

This integration process leads to the students being able to establish new fast routes between the clusters and the concepts within those clusters, improving knowledge retrieval efficiency and allowing the students to make use of concepts to reason through the clinical cases. As a consequence of these new fast routes into the

medical knowledge map it is expected that students would start to use more elaborate concepts to discuss clinical cases, for example a serious generalised infection with high concentration of bacteria in the blood becomes a *sepsis.* The words sepsis in an expert's mind would also be associated with specific symptoms, affected organs, biological processes taking place, clinical investigations, treatments and other biomedical knowledge. As Boshuizen describes *"a new type of clinical or semi-clinical concept appears in the protocols, such as micro-embolism, aorta-insufficiency, forward failure, or extra-hepatic icterus, providing a powerful reasoning tool"* (Boshuizen, 2003 p. 12).

Studying students' use of language is a useful way of understand this process, as the encapsulation of knowledge should be reflected in students use of technical language. As students progress through the years of PBL discussions their language should become increasingly more elaborated, and they should use less words to refer to more complex concepts and engage in more biomedical-clinically focused discussions.

The results found through electronic content analysis, identified differences in the language used by the two groups that seem to confirm our hypothesis of knowledge encapsulation. In order to better understand those differences we used corpus analysis to conduct further research into these differences.

All the discussions recorded that first-year students are using significantly ($p<0.0001$) more medical terms (common and technical) then the ones used in a normal discussion in English (Table 5. 9 reports the mean by case of common and technical medical words for first-year group by module, week in the study and case. It is possible to observe that there is a wide variability of use of technical words each week, with a slight increase on the frequencies towards the final weeks in each of the modules.

Relative frequency of words/session allows for comparisons to be made independently of the total number of words in the session/case transcript. It is important to notice that as in any other spoken text there are types of syntactic and semantic elements in the texts. The tag set used for Wmatrix2 has 135 tags plus 12 for punctuation therefore it is not expected to find relative frequencies of single categories higher then 10%, this value representing an extremely frequent occurrence (Rayson 2008). Table 5. 10 presents the descriptive values for each case, minimum and maximum frequencies and variance of medical words in each case

(Table 5. 8). This was achieved by comparing all the transcripts in the session with the sample of British National Corpora Spoken text. This sample, built during the last decade of research on the uses of British English by the British National Corpora consortium, was used as a baseline measure to ensure that common and technical[36] medical terms use in the PBL discussions were higher then in a normal English conversation. As expected, the log-likelihood values are extremely high, especially in the technical words/expressions, as these are not part of common speech, but rather part of discourse in the medical field.

Table 5. 9 reports the mean by case of common and technical medical words for first-year group by module, week in the study and case. It is possible to observe that there is a wide variability of use of technical words each week, with a slight increase on the frequencies towards the final weeks in each of the modules.

Relative frequency of words/session allows for comparisons to be made independently of the total number of words in the session/case transcript. It is important to notice that as in any other spoken text there are types of syntactic and semantic elements in the texts. The tag set used for Wmatrix2 has 135 tags plus 12 for punctuation therefore it is not expected to find relative frequencies of single categories higher then 10%, this value representing an extremely frequent occurrence (P Rayson 2008). Table 5. 10 presents the descriptive values for each case, minimum and maximum frequencies and variance of medical words in each case.

Table 5. 8: Comparison between first-year discussion and BNC frequency of medical words/expressions

Relative frequency of Technical words/expressions TMW (Average)	LL vs. Normadata (BNC spoken sample)	Relative frequency of common medical words/expressions CMW (Average)	LL vs. Normadata (BNC spoken sample)
4.51	(+) 44734.85***	4.04	(+) 12938.44***

Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category). LL= Loglikelihood * P < 0.01; critical value = 6.63 **P < 0.001; critical value = 10.83 ***P < 0.0001; critical value = 15.13

36. For the purposes of this research we differentiate between common and technical medical words/terms. Common medical terms (CMW) being the ones used by lay and experts in medicine equally, such terms are disease, cold, pain, doctor, clinician, X-ray. Technical terms (TMW) are specialised words used in medicine, part of a special Medical Lexicon, that is used by professionals in the field (medical jargon), example nasopharyngitis, rhinopharyngitis as synonymous of a common cold.

Table 5. 9: First-year mean of common and technical medical words by case

Module	Weeks of study	Cases	Relative Frequency of common medical words/expressions *	Relative Frequency of Technical words/expressions *
Respiratory	Week 1	2.2-To cap it all	3.79	4.18
	Week 2	2.3 It's affecting my football	3.85	4.13
	Week 3	2.4 Can we fix it?	4.26	4.38
	Week 4	2.5 Out of Africa	5.00	4.50
Cardiovascular	Week 5	3.1 Life is unpredictable	3.23	3.75
	Week 6	3.2 Be still my beating heart	3.57	4.28
	Week 7	3.3 It can wait until Monday	3.98	4.89
	Week 8	3.4 Old and crusty	3.60	5.09
	Week 9	3.5 Benign or not benign?	4.97	5.31

*Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category).

Table 5. 10: First-year distribution of common and technical medical words by case

	Relative frequency of common medical words			Relative frequency of technical words		
Clinical case	Max	Min	Variance	Maxi	Min	Variance
2.2 To cap it all	4.15	3.42	.27	4.38	3.98	.08
2.3 It's affecting my football	4.19	3.40	.16	4.76	3.51	.39
2.4 Can we fix it?	4.84	3.45	.52	4.71	4.07	.10
2.5 Out of Africa	6.28	4.31	1.23	5.15	3.82	.44
3.1 Life is unpredictable	3.50	3.07	.05	4.49	3.12	.48
3.2 Be still my beating heart	5.24	2.47	2.17	5.70	2.79	2.12

3.3 It can wait until Monday	4.88	3.39	.62	5.01	4.81	.01
3.4 Old and crusty	4.35	2.49	.96	5.75	4.63	.34
3.5 Benign or not benign?	5.23	4.65	.09	5.82	5.04	.20

*Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category). a. year of study = 1

Table 5. 10 shows a wide range of variance values between the sessions in each case, with a minimum variance for technical terms being observed in the first case 2.2-To cap it all and the highest on case 3.2, Be still my beating heart. The relative frequency of common terms is equally variable with a minimum of 0.5 in the 3.1, Life is unpredictable, case and a maximum of 2.17 observed in 3.2, Be still my beating heart. Further statistical analysis leads to the conclusion that no significant difference between individual cases or modules were found for the year 1 group (Table 5. 11 and Table 5. 12).

It is possible to see that there is an increase on the frequency of technical words used by year one students as they progress over time through the different cases. This is confirmed by a significant positive correlation between the clinical case number and technical words shown in Table 5. 13. Additionally, we tested for significant correlations between the number of technical and common medical words and the position of the session in each of the modules. A positive significant correlation would confirm that the use of technical words would increase in a linear shape as students acquire knowledge and progress through the module. In the respiratory module there is a significant positive correlation showing an increase of the use of common medical words as the students' progress through the discussions within the module. Interestingly in the cardiovascular module are the technical medical words that register such increase over time (Table 5. 14).

Table 5. 11: Statistical comparison between sessions (Kruskal-Wallis Test)

Clinical case	Relative frequency of common medical words	Relative frequency of Technical words
	Mean Rank	Mean Rank
2.2-To cap it all	11	8.5
2.3 It's affecting my football	11.5	9
2.4 Can we fix it?	16.33	10.5
2.5 Out of Africa	20	13.67
3.1 Life is unpredictable	6.33	6
3.2 Be still my beating heart	9.67	11.5
3.3 It can wait until Monday	13	18.33
3.4 Old and crusty	10.5	19.33
3.5 Benign or not benign?	22.33	23
Test Statistics(a)(,)(b)(,)(c)		
	Relative frequency common medical words	Relative frequency of Technical words
Chi-Square (H(2))	10.864	13.022
Df	8	8
p value	0.21	0.111

year of study = 1 b. Kruskal Wallis Test c. Grouping Variable: Clinical Case

Table 5. 12: Statistical comparison between modules (Kruskal-Wallis Test)

Ranks(a)	Relative frequency of common medical words	Relative frequency of Technical words
Module	Mean Rank	Mean Rank
Respiratory	15.05	10.59
Cardiovascular	12.37	15.63
Test Statistics(a)(,)(b)(,)(c)		
	Relative frequency of common medical words	Relative frequency of Technical words
Chi-Square	0.779	2.759
Df	1	1
p value	0.378 (NA)	0.097 (NA)

year of study = 1b. Kruskal Wallis Test c. Grouping Variable: Module

Table 5. 13: Correlation between clinical case and technical and common words (Spearman's) Year 1

	Spearman's correlation	% total number of common medical words	% total frequency of Technical words
Clinical case number	Correlation Coefficient	0.134	.590**
	p value (2-tailed)	0.513	0.002
	N	26	26

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5. 14: Correlation between session in each module and number of technical and common medical words

	Correlations	Relative frequency of common medical words	Relative frequency of Technical words
Respiratory	Spearman's Correlation	.618*	.355
	Sig. (2-tailed)	.043	.285
	N	11	11
Cardiovascular	Spearman's Correlation	.450	.732**
	Sig. (2-tailed)	.092	.002
	N	15	15

**Correlation is significant at the 0.01 level (2-tailed).

Second-year discussions, as expected, showed a significantly higher frequency of medical (technical and common) words/expression then expected in a conversation in English, as shown by the high LL values (Table 5. 15). Table 5. 16 presents the means of technical and common medical words or expressions by module/week/case and Table 5. 17 presents the descriptive values for each case, minimum and maximum frequencies and variance of medical words in each case.

Table 5. 15: Comparison between second-year discussion and BNC frequency of medical words/expressions

	LL vs. Normadata (BNC spoken sample)
Relative frequency (%) of common medical words/expressions (B)	+ 5981.97***
Relative frequency (%) of technical words (V)	+ 30044.67***

Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category). LL= Loglikelihood * P < 0.01; critical value = 6.63 **P < 0.001; critical value = 10.83 ***P < 0.0001; critical value = 15.13

Table 5. 16: Second-year means of common and technical medical words by case

Module	Weeks of study	Cases	Relative frequency of common medical words/expressions (B) (Average)*	Relative frequency of technical words (V) (Average)*
Neurosciences	Week 1	8.4 Only the lonely	4.05	5.29
Neurosciences	Week 2	8.5 Twisted Sister	4.58	5.68
Neurosciences	Week 3	8.6 Shaky Nerve	2.90	7.34
Integrative	Week 4	9.1 Hard labour	4.03	4.50
Integrative	Week 5	9.2 Blood brothers	6.15	5.14
Integrative	Week 6	9.3 More than just a rash	2.46	4.70
Integrative	Week 7	9.4 Put your foot down	3.72	4.26

*Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category).

Table 5. 17: Second-year distribution of common and technical medical words by case

	Relative frequency of common medical words (B)			Relative frequency of technical words (V)		
Case	Maximum	Minimum	Variance	Maximum	Minimum	Variance
8.4 Only the lonely	5.12	3.04	1.08	5.92	4.74	.35
8.5 Twisted Sister	4.79	4.36	.09	5.69	5.67	.00
8.6 Shaky Nerve	2.90	2.90	.**	7.34	7.34	.**

9.1 Hard labour	4.43	3.77	.12	5.25	4.01	.43
9.2 Blood brothers	6.15	6.15	.**	5.14	5.14	.**
9.3 More than just a rash	3.11	1.98	.34	7.00	2.89	4.40
9.4 Put your foot down	4.98	2.46	3.18	5.08	3.43	1.36

*Relative frequency (%) is calculated by (absolute number of words in category/total number of words in a text)*100. This measure describes the % of text dedicated to common medical words (B category) and to technical terms (V category).** Only one session recorded in this case.

On the neurosciences module, the frequency of technical terms increases from the first to the last recorded case in the module. However, that is not the case on the integrative module. Variance between sessions in each of the recorded cases reveals, as for the first-year group, a wide range of values. A maximum variance identified of 4.40 on case 9.3, More than just a rash, and a minimum of approximately 0 in the case 8.5, Twisted sister, indicating that the frequency of technical words/expressions was almost the same across the sessions of this case. Regarding common medical words/expressions, the same phenomenon was identified with the last case recorded 9.4, Put your foot down, presenting a variance of 3.18 and the 8.5, Twisted sister, being the case where the frequency of these words/expressions was more consistent across sessions with a variance only of 0.09 between sessions.

For this group no significant differences can be observed between the cases, or between modules when considering common medical words Table 5. 18. Although technical words are (statistically) significantly more frequent in the Neurosciences modules with a mean rank of 11 then in the Integrative module with a mean rank of 6 ($H(1)=4.5$ $p=0.034$). As with first-year modules, we tested for a linear increase of the number of technical and/or common words within each module; however, in this group neither shown significant correlations (Table 5. 20).

Table 5. 18: Statistical comparison between sessions (Kruskal-Wallis Test H)

Clinical case	Relative frequency of common medical words	Relative frequency of Technical words
	Mean Rank	Mean Rank
8.4 Only the Lonely	9.33	9.33
8.5 Twisted Sister	11	11.5
8.6 Shaky Nerve	4	15
9.1 Hard labour	8.67	6
9.2 Blood brothers	15	8
9.3 More than just a rash	3	6.33
9.4 Put your foot down	8	4.5
	Test Statistics(a)(,)(b)(,)(c)	
	Relative frequency of common medical words	Relative frequency of Technical words
Chi-Square	8.233	6.183
Df	6	6
p value	0.222	0.403

year of study = 2 b. Kruskal Wallis Test c. Grouping Variable: Clinical case

Table 5. 19: Statistical comparison between modules (Kruskal-Wallis Test)

Ranks(a)	% total number of common medical words	% total frequency of Technical words
Module	Mean Rank	Mean Rank
Neurosciences	9	11
Integrative	7.33	6
Test Statistics(a)(b)(c)		
	Relative frequency of common medical words	Relative frequency of Technical words
Chi-Square	0.5	4.5
Df	1	1
p value	0.48	0.034*

year of study = 2 b. Kruskal Wallis Test c. Grouping Variable: Module

Table 5. 20: Correlation between session in each module and number of technical and common medical words

	Correlations	Relative frequency of common medical words	Relative frequency of Technical words
Neurosciences	Spearman's Correlation	-.029	.371
	Sig. (2-tailed)	.957	.468
	N	6	6
Integrative	Spearman's Correlation	-.283	-.083
	Sig. (2-tailed)	.460	.831
	N	9	9

No significant correlation.

Comparing the frequency of medical terms by session no single clear pattern emerges from both years. On the first-year group the highest frequency of common medical terms was identified in the first sessions, while in the second group it is the second session that shows the highest frequency for this category. The highest frequency of technical terms for the first-year group is observed in the third and last session of the case, while for the second-year group it is the second session which has the highest frequency. Nevertheless, overall no (statistical) significant differences in the frequencies of common or technical words/expressions were found between sessions within years or for both years combined, leading to conclude that there is a similar use of these words/expressions across all three sessions in both years.

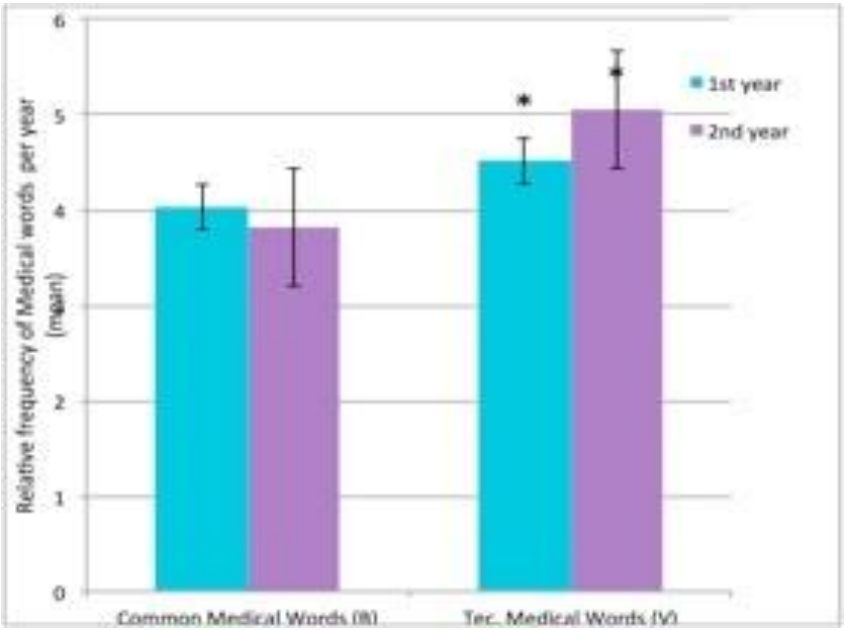
Table 5. 21: First and second-year distribution of common and technical medical words by session

Sessions	Year 1		Year 2	
	Relative frequency of common medical words (B)	Relative frequency of Technical words (V)	Relative frequency of common medical words (B)	Relative frequency of Technical words (V)
1	4.23	4.35	3.04	4.27
2	3.96	4.47	4.11	5.35
3	3.94	4.71	3.88	5.11

Figure 5. 4 show the relative frequency of common and technical words/expressions between years of study. As reported in the previous results the frequency of technical terms is higher than the frequency of common medical terms. This was an expected result due to the context and nature of the PBL sessions.

Second-year students use (statistically) significantly more technical words in their discussions than first-year students corroborating the initial hypotheses of knowledge encapsulation being detected via analysis of the PBL sessions. Technical words are (statistically) significantly more frequent in the second-year group for a significance level of p=0.045 as calculated by Kolmogorov-Smirnov rejecting the null hypothesis of no differences in the distributions. With regards to common medical words, although first-year students use more of these words, no significant difference is registered.

Figure 5. 4: First and second-year distribution of common and technical medical words



* Statistical significant difference for p<0.05 (Kolmogorov-Smirnov two independent samples)

The table below shows the subcategories of technical words presenting the highest (statistical significant) differences, as calculated by the Log-likelihood (LL) for a p<0.01 and a critical value higher then 6.63. Interestingly, the subcategories related to treatments and surgical procedures are the ones where the highest differences are found. Also of note is the differences found in the diagnosis subcategory, implying that the second-year students use more often words/expressions related to

either identifying a diagnosis (e.g. Cancer, glossopharyngeal neuralgia) or key aspects that lead to its definition (e.g. haemolytic, ketoacidosis).

Finally, although it would be expectable, with some indications in the data that variations in relative frequency of technical and common words would be related, our results show no significant correlation between the use of common and technical words (Spearman correlation of 0.231 p=0.147) overall. This does not show that there is no association or relationship, but rather no linear dependence between these two categories shown in our data.

Table 5. 22: Differences in subcategories of technical language between years

Technical language subcategory	Relative frequency year 1	Relative frequency year 2	Direction*	LL**
Surgical procedure	0.13	0.32	-	136.81
Treatment by medication	0.69	1.1	-	133.57
Chemicals	0.01	0.08	-	127.61
Descriptive information	0.09	0.2	-	57.34
Medical devices	0.02	0.07	-	41.63
Amount or degree (related to tests)	0.09	0.16	-	28.19
Patient record terms	0	0.02	-	20.46
Indication of change (related to tests)	0	0.01	-	18.4
Diagnosis	0.56	0.67	-	14.18
Hospitals, clinics, physicians, staff	0.02	0.04	-	11.49
Words referring to patient	0	0.01	-	10.57
Verbs of observation	0.01	0.02	-	7.79
Disease indicator word	0.42	0.49	-	6.69

*Direction: + indicates overuse in year1 relative to year2, - indicates overuse by year 2 relative to year1
**LL significance values: P < 0.01; critical value = 6.63 ; P < 0.001; critical value = 10.83 ; P < 0.0001; critical value = 15.13. The high the LL value is the large is the difference between the compared sessions.

Questions

Questioning is an essential skill of the medical profession since the questions that a doctor asks a patient can lead to the necessary information that will help guide their thought and reasoning processes in order to make a diagnosis. The ability to ask questions in order to get appropriate information from the patient's answers, is a cornerstone of effective clinical reasoning (Walker, 2003; Visschers-Pleijers *et al.*, 2006)

The literature on expertise in medicine and clinical reasoning highlights differences between novices and experts in these skills (Chapter II: Background). While novices tend to ask more questions and have a more disorganised approach to patient inquiry, experts ask fewer questions which are more directed to confirm or refute diagnostic hypotheses formed from other elements of information about the patient, for example risk factors, age, gender, physical appearance, presentation or description of main complaints (Kassirer *et al.* 2009; Patel *et al.* 2005). The PBL cases are designed to help students develop these skills (H.S Barrows & Tamblyn 1980; Goss *et al.* 2011); however, contrary to other learning situations, for example clinical skills sessions dedicated to history taking where simulated patients or actors are role playing patients, in PBL sessions questions are not directed at patients.

Here the cases begin with a patient presentation, usually delivered to the students by a video of the patient, followed by additional information given by the facilitator. The students' task is to interrogate the facilitators, colleagues, and other knowledge sources in order to gather the information they need to learn in order to progress through the clinical case. Questions that the group would like to ask the patient are highlighted as part of the learning points in the case. Questioning is also of course a crucial part of any active discussion, especially if the discussion involves challenging other’s ideas or asking for clarification of concepts. From this point of view one can consider the frequency of question as a measure of how prolific the discussion taking place is. Question marks were the main identifiers of questions. During the validation study we concluded that question marks were the best element to identify questions, especially when compared with "question words". The latter had been defined as words that are typically associated with questions, such as what, why, how, where, when or which. However, since as they are also found in non-questioning speech as well as in questions, using them as questions identifiers would be misleading.

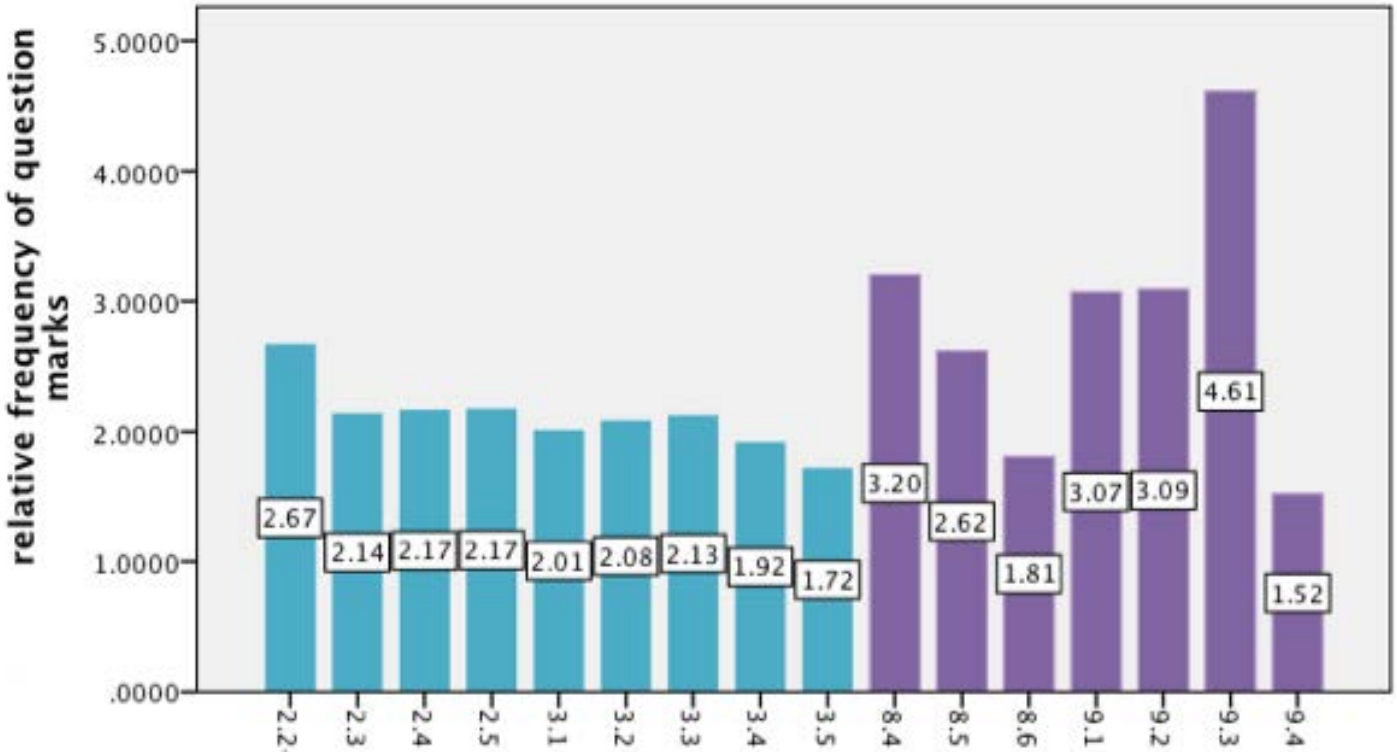
Table 5. 23: Example of questioning concordances

Easy one thromboxane A2, and this receptor, what was it, do you know?
Does that picture means that it acts on thromboxane to receptor…?
Then there's xiii to turn fibrin into cross-linked fibrin?
Yeah,... so you'll find the common pathways a bit easier, start off with what?
Obviously the PaCO2 would be normal before you compensation, what's it detecting?
... it's called milk-alkali syndrome, have you heard of that ?
Even though it's metabolic alkalosis it could be metabolic compensation as well?

Figure 5. 5 shows the average of relative frequency per case analysed for both first and second-year groups. It is possible to notice that, five of the seven second-year cases revealed higher use of questions than the first-years' cases and while first level of questions seems to be quite constant, as time progresses there is higher variability across cases for the second-year discussions. Also, case 2.2 for the first-year and cases 8.4 and 9.3 seem to be the ones generating more questions for the respective groups. However, no significant difference in the use of question per case was found for the first or the second-year group. Figure 5. 6 presents the mean of questions per session for both groups.

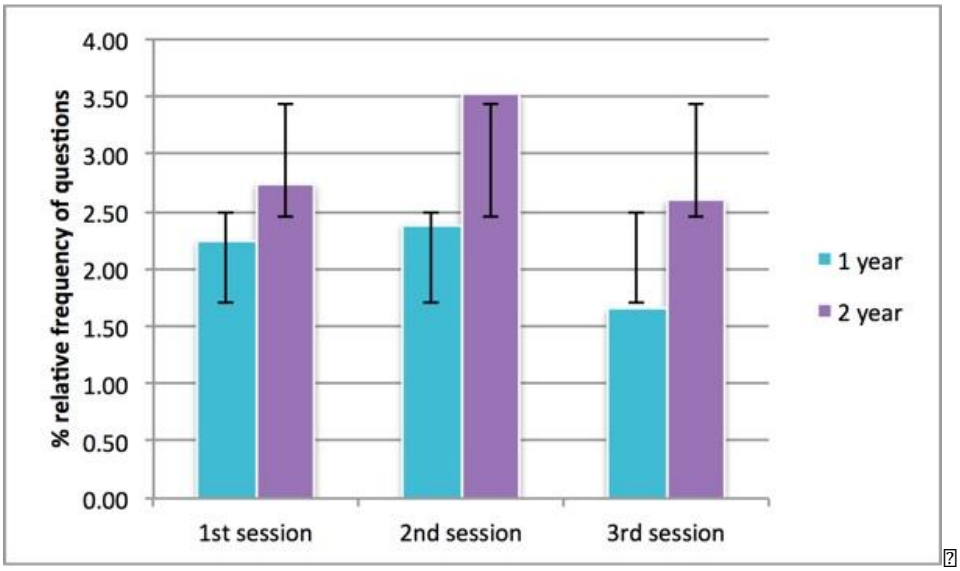
Interestingly for both groups the second session is the one reporting the highest frequency of questions, although these are not statistically significant differences. Interestingly, the data contradicts our initial hypothesis that the second-year students would ask more questions in the first sessions, as it is the session where students are for the first time presented with the patient information plus this is where it is expected that the students will be defining their learning aims for the session. However, a similar result was found when comparing the use of question per year, that is the total use of questions by the second-year is clearly higher than first-year, but this difference is not significant for the Kolmogorov-Smirnov test statistic of 1.313 a significance level (2-side test) of 0.64 (Figure 5. 7).

Figure 5. 5: Distribution of the (mean) of questions per case for the first and second-year groups



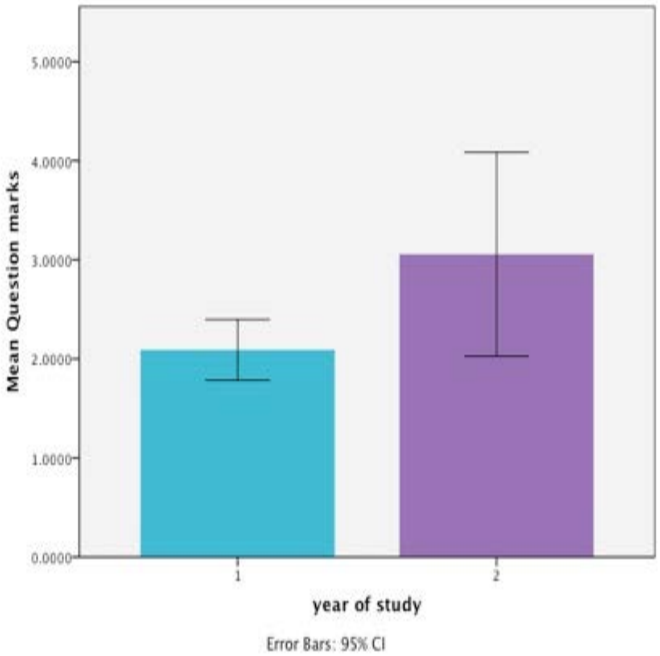
Blue=first-year data Purple=second-year data

Figure 5. 6: Mean of questions per session type and year groups (error bars 95%)



Blue=first-year data Purple=second-year data

Figure 5. 7: Mean of questions per year (error bars 95%)



Blue=first-year data Purple=second-year dat

Explanations and Reasoning

Explanations were measured by the semantic categories A2 and A7. These are large subsets of the UCAS system based on the words identified through many decades of research into British language use in conversations (spoken BNC).

The semantic category A2 encompasses words that: demonstrate change; words that express modification; and words that express cause, effect and connection. A7 groups the words or multi-word expressions that express ideas related with probability and likelihood. To further identify reasoning utterances we also used subordinating conjunctions (CS). These are words such as 'if…so', 'then', 'when' and 'because', used in English to join a subordinate (dependent) clause to a main (independent) clause. Subordinating conjunctions occur when individuals are contrasting, comparing, relating and connecting and are useful markers for reasoning, explanatory and causal sentences. By the analysis of the frequency of these categories, it is possible to identify explanation and reasoning utterances. However, as was pointed out before, the validation study led to the conclusion that it is difficult to distinguish clearly between explanation utterances and the utterances of reasoning alone. Therefore the results presented here are for both reasoning and explanations analysis combined.

Table 5. 24: Example of Explanations and reasoning concordances

Extrinsic allergic alveolitis	**although**	I'm not sure **because of**
you've got a small VSD it can get bigger or	**whether**	it's just been misdiagnosed.
I just said it could go up with allergies	**because**	it is an allergy.
The living conditions could	**cause**	all this .
It does	**change**	lots of parameters,
He	**could**	have a congenital abnormality like DVA
It	**could be**	the fact that I've got here Downs Syndrome
	Maybe	he has got double pneumonia.
of HCO3 but I don't know what would be the	**reason**	for that. But I **suppose** you could be exposed
you've got things like (…) artery pressure,	**so what**	we need to do is get an ECG, get an Echo…
abuse of antacid **so** you **could** have those I	**suppose**	.
so plaque kind of breaks and	**then**	an aneurysm forms one or the other **could** compress.
I	**think**	they are actually considered part of the basal ganglia

Table 5. 25: Relative frequency of explanations and reasoning per case for first and second-year

Year	Module	Case (cycle)*	Relative frequency (%) of explanations and reasoning per case
1	Respiratory	2.2-To cap it all	11.73
1	Respiratory	2.3 It's affecting my football	18.65
1	Respiratory	2.4 Can we fix it?	19.78
1	Respiratory	2.5 Out of Africa	19.36
1	Cardiovascular	3.1 Life is unpredictable	21.94
1	Cardiovascular	3.2 Be still my beating heart	20.17
1	Cardiovascular	3.3 It can wait until Monday	22.39
1	Cardiovascular	3.4 Old and crusty	18.87
1	Cardiovascular	3.5 Benign or not benign?	19.55
2	Neurosciences	8.4 Only the lonely	14.6
2	Neurosciences	8.5 Twisted Sister	13.8
2	Neurosciences	8.6 Shaky Nerve	5.5
2	Integrative	9.1 Hard labour	17.58
2	Integrative	9.2 Blood brothers	7.07
2	Integrative	9.3 More than just a rash	16
2	Integrative	9.4 Put your foot down	12.62

The relative frequency of explanations and reasoning are higher for the first-year cases, with a maximum of 12.32 for 3.1, Life is unpredictable, case and a maximum of reasoning for 3.3, It can wait until Monday, 10.64. For the second-year group both explanations and reasoning are higher for the case 9.1, Hard labour, 8.86 and 8.72 respectively. Although these differences can be identified, they are small and not statistically significant with a $H(8)=9.770$ $p = 0.282$ for the first-year group and $H(6)= 6.720$, $p = 0.347$ for the second-year group.

Nevertheless, the first-year group, seems to be engaging in significantly more explanations and reasoning in their second module, the cardiovascular module, than in their respiratory module with a mean rank of 16.30 versus a mean rank of 9.68 ($H(2)=4.753$ $p=0.029$) respectively.

For the second-year, although the second module (Integrative) reports a slightly higher (mean rank of 9.44) when compared with the neurosciences module (mean rank of 5.83) that difference is not statically significant for a $p<0.05$ ($H(2)=2.351$, $p=0.125$).

Table 5. 26: Relative frequency of explanations and reasoning per sessions of the PBL cases

PBL session	1 year	2 year
	Explanations & Reasoning	Explanations & Reasoning
First session	6.85	4.89
Second session	6.66	5.79
Third session	6.41	6.39

No significant differences reported (Kruskal-Wallis Test H).

An analysis of the patterns of the frequencies of explanations and reasoning per sessions shows that these utterances are more frequent in the first session for the first-year group, while for second-year students it is in the last session (third session) that the highest frequency occurs. Additionally, differences between years are also found when analysing the explanations and reasoning utterances individually.

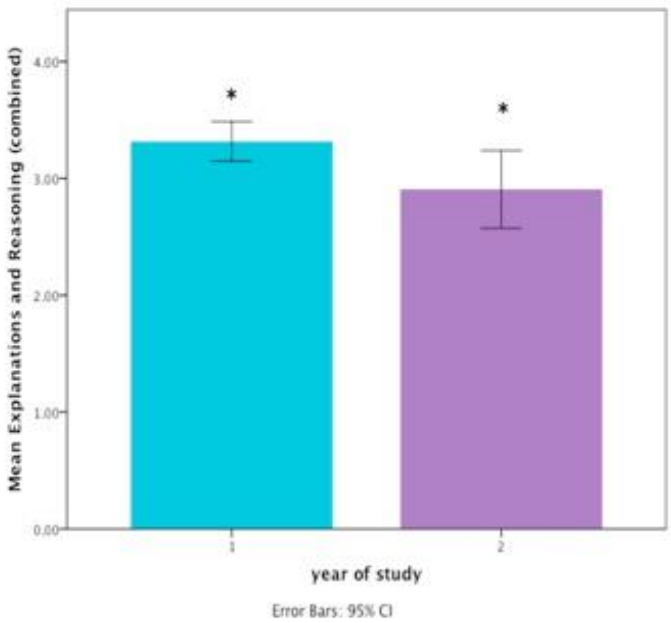
First-year students seem to engage more frequently in explanations on the first session with reasoning utterance being more frequent in the second session. For the second-year, it is the third session that captures the highest frequency of utterances on both explanations and reasoning. These differences are, however, of small size and are not statistically significant, therefore, there seems to be no significant impact of the session (first, second or third) on frequency of explanations and reasoning.

Table 5. 27: Statistical comparison between sessions (Kruskal-Wallis Test H)

Explanations and Reasoning (combined)				
Groups	first-year		second-year	
session of the PBL case	N	Mean Rank	N	Mean Rank
First session	8	16.38	4	7
Second session	9	13.56	6	7.08
Third session	9	10.89	5	9.9
Total	26		15	
Test Statistics (a,b)				
Chi-Square	2.18		1.357	
Df	2		2	
p value	0.336 (NA)		0.507 (NA)	

a. Kruskal Wallis Test b. Grouping Variable: session of the PBL case No statistical significant differences

Figure 5. 8: Mean of explanations and reasoning (combined) year1 vs. year2



* Statistical significant difference for p<0.05 (Mann-U)

Table 5. 28: Explanations and Reasoning (combined) statistical comparison between years (Mann-Whitney U)

	Explanations and Reasoning (combined)		
Year of study	N	Mean Rank	Sum of Ranks
1	26	24.77	644
2	15	14.47	217
Total	41		
Test Statistics(b)			
Mann-Whitney U	97		
Wilcoxon W	217		
Z	-2.653		
p value (2-tailed)	0.008		
p value [2*(1-tailed Sig.)]	.007a		

a. Not corrected for ties. b. Grouping Variable: year of study

Analysing the difference between years, it is possible to observe that explanations and reasoning utterances are significantly more frequent in first-year students' discussions with a mean rank of 24.77 (N=26), than in the second-year discussions with a mean rank of 14.47 (N=15) and a Mann-Whitney U=97.000 for a p=0.007 (2-tailed).

B. Limitations

In order to be able to discuss the results presented in the last section, it is necessary to reflect on a few of the limitations of our study. These do not undermine the work carried out or the results found; however, their identification is an important part of any investigation and a very important step in guaranteeing the truthfulness of the analysis and discussion of the results.

Tension between qualitative and quantitative data

The present study is a mixed-methods study, using type III approach to the analysis (Paul Rayson 2008). Qualitative and quantitative views of the data are combined in the iterative process of achieving understanding of the data with a focus on the quantitative presentation of data. Creswell and others identify the tension between qualitative and quantitative methods as one of the issues inherent to mixed-methods approach (Creswell et al. 2003; Creswell & Clark 2007; Zhang 2011; Heyvaert et al. 2011; Creswell & Clark 2011) that should be addressed by the researchers. These tensions were not alien to the present research, often decisions had to be made based on the readings of the materials to correct or change the quantitative outputs, for example, when utterances included in certain categories were not correctly allocated. An activity log was kept to ensure traceability and audit of all decisions was possible.

The iterations from quantitative view to a qualitative view give the researcher a deep relationship with the data, this can lead to the interest in searching the data for new patterns, for new interesting arising aspects. This needs to be done carefully, as it can create an increase in the research bias, rather then a decrease advocated by these methodologies. To avoid this, part formed the mentioned diary of the research, where the main analysis steps were registered. Also we focused our analysis in the research question and aims defined at the beginning of our data analysis.

A related aspect is the tension felt in the reporting of the results. Traditions from qualitative and quantitative perspectives demand very different types of data reporting, not easy to combine in a coherent narrative. Our aim was to be able to identify patterns, tendencies in the data for that reason quantitative results would be the most appropriate. Hence our decision to focus results on the quantitative results using adequate text extracts followed the example from others in healthcare

linguistics (Crawford & Brown 2010; Friedman *et al.* 2009; Gooberman-Hill *et al.* 2009; Skelton & Hobbs 1999; Skelton *et al.* 2002; Thomas & Wilson 1996; Harvey *et al.* 2007).

Data sources

Our study followed only two groups of volunteers; this should be taken in consideration when interpreting the results. For corpus analysis the sample sizes are defined not by the number of sources but by the size of the corpus collected (60 hours of discussions in our study) in relationship with the aim of the study and how well that corpus captures what is being studied. Both size and scope of our corpus are aligned with what others have advocated for healthcare and healthcare education corpora (P Crawford & B Brown 2010; Friedman et al. 2009; Brian Brown et al. 2006). Although our corpus included only two groups one in each different year, the corpus collected involved discussions with several different cases, within four total modules, and involving two different facilitators per group (one clinical and one non-clinical). This is adequate to our study aim, which was not to generalise findings across the population of medical students or PBL students. The aim was to understand how in the early stages of a PBL curriculum students without extensive medical knowledge approach clinical cases, and how those compared with students that had already gained a more comprehensive knowledge base (second-year students). For this purpose, studying two groups over a longer period of time was deemed the more appropriate strategy.

Individual differences

Both groups had a similar distribution of gender, age, and background (biomedical vs. non-biomedical) as shown in the results section. These are critical elements, as for example a higher number of students with a biomedical background in one group could lead to an increase of the technical vocabulary used, or increase the level of explanation and reasoning, as these students would already have more established networks of biomedical knowledge they can easily mobilise to new problems. Also, these students could already have developed clinical reasoning strategies by their contact with clinical practice and that would skew our data towards a group with a

high number of these students. However, the distribution proved to be quite similar. This is supported by the average of students' results on GAMAST, which is used as a measure of the students' minimum required knowledge on humanities and social sciences, biological and physical sciences was similar between two groups. We do recognise that, other individual differences such as personality, may influence the way the students interact or engage with the group and/or the PBL process and the dynamic of each group may have an impact in the particular student participation. However, as Berlin states *"Educational researchers have to accept the embeddedness of educational phenomena in social life, which results in the myriad interactions that complicate our science."* (Berliner, 2002). Masking this reality is neither possible nor desirable therefore we do not attempt to make generalisations that go beyond the context of this research.

Transcription errors

Transcriptions were conducted by a professional with experience in medical language and checked against the video files by the researcher. However, due to the challenge of recording seven students and a facilitator discussing and sharing ideas, some errors were still detected in the transcripts. Three main error sources were identified and corrected by the researcher. One when the recording was not clear enough to be transcribed, a second when participants were talking over each other in such a way that it was not possible to clearly identify one of them. A final source of error in the transcripts was the misspelling, in a few occasions of some medical terms.

Unequal number of sessions

The number of sessions recorded for the two groups was not equal, with more sessions being recorded in the first-year. This was due to a delay in the recruitment process for the second-year group caused by external reasons plus rearrangements of the times of the sessions on a few occasions. However, it was possible to ensure that PBL cases recorded were diverse and covered two modules as with first-year students, not affecting the balance or comparability of the corpora. Additionally as all the statistical measures were calculated based on relative frequencies, a common practice in corpus analysis, the differences in absolute numbers do not undermine the results, nevertheless, this needs to be considered when discussing the results.

Multidimensionality of human communication

Human communication is multidimensional and is an extremely rich source of data. Language is just a part of a complex communication system that we use to share meanings. Corpus analysis is a method of analysing language by analysing the semantic and syntactic value of basic observable units of meaning, that is, words. This method is not suited to studying tacit meanings or hidden messages in communication, rather it suits studies, such as ours, that aims to study the explicit use of language in particular contexts and draws conclusions on that basis.

C. Conclusions

Our study was designed to investigate how students within two PBL groups approach clinical cases and what are the main topics emerging form these discussions? How are the students making use of technical language? And what are the strategies the students are using to discuss these cases, mainly looking at variations and patterns in questions and explanations and reasoning. However, it is important to note we set out this study to investigate the explicit and verbalised cognitive processes not internal. The study of these would require different methodologies.

Aim 1: Understand the use of technical language in exploration of the clinical cases

The knowledge encapsulation hypothesis states that students start to build their networks of knowledge by clustering detailed biomedical concepts together, based on their connections under higher level technical concepts (Boshuizen *et al.,* 1992). For example, a general infection in the body would then become clustered into *sepsis*, under this high level concept about the physiopathological mechanisms, probable causes and consequences would be a collection in the memory (Boshuizen *et al.,* 1992). Although there are some criticisms of this concept, this is still one of the most commonly used explanations for the development of clinical reasoning in medical students (Boshuizen, 2003), and hence is a valid hypothesis to be tested within our data set.

Encapsulation would mean for our study that, when comparing the first and second-year groups, second-year students would show an increase in the use of technical language (*medical jargon*) and a decrease in the use of common medical vocabulary. As concepts would be clustered into more high level technical concepts, it would be hypothesised that the overall number of words would decrease, as students would not need to use such detailed step-by-step explanations to communicate their ideas about the case. Reasoning would still follow a step-by-step approach but steps considered would be larger (Boshuizen, 2003).

Also differences would be expected to occur between cases within the modules/cycles. Modules/cycles were divides by systems with all cases addressing particular aspects of that system. As the module progresses the students would gain

more knowledge regarding that system, therefore an increase in knowledge encapsulation would be expected.

Our data also shows that in PBL conversations the relationship between technical and what was defined as common medical words is more complex than initially suggested. Our data showed that in fact no direct correlation between these two semantic categories could be found. These discussions are highly technical, but they are also informal discussions and learning environments where high levels of colloquialism are to be expected. However, data confirms that in fact the language used in the PBL sessions is significantly different from a typical conversation in English, even for the first-year students still in their very early days into the curriculum. If this was motivated by the biomedical students in the group (3 in 7) that are leading to specific points in the transcripts where high frequencies would be detected. However our data does not support this hypothesis as technical vocabulary is spread out through the transcripts and discussion. Nevertheless our analysis does not allow the determination of exactly what percentage of that is due to the students with biomedical background interventions.

In fact, as hypothesised, the density per hour of the first-year group is more than two thousand words higher almost then second-year group (from 4156 w/h to 6115 w/h). Personality and other confounding variables such as group dynamics can have an impact on this, however, bearing in mind the similar group constitution of the groups in gender, age, backgrounds and GAMSAT results the hypothesis that this result maybe due to the fact that the group is in an more advanced stage of their learning should also be considered.

Electronic content analysis (ECA) also shows that second-year students use more technical concepts (medical lexicon) and first-year students seem to be discussing more aspects related with the case information. This was confirmed by corpus analysis with second-year students using significantly ($p<0.05$ Kolmogorov-Smirnov) more technical and common vocabulary than first-years. Nonetheless, the emergence of ‘themes’ such as ‘things’, ’stuff’ or ‘ones’ is noticeable. This expresses both colloquialism of the conversation and the fact that although using more technical words, these students are still in the process of constructing their medical vocabulary, which would at this stage be expected.

No significant differences were found from the comparison between first session vs. second session vs. third session in neither years analysed. Students’ use of technical

language seems to be consistent across all types sessions in each of the years analysed.

The second-years' the first module, (8. Neurosciences) showed an increase with time of the relative frequency of technical words and a decrease in common words. However, this was not the case on the second module recorded (9. Integrative). Here a decrease in both categories over time was observed (although not statistically significant). Contrary to what would be expected the use of technical terms was significantly higher for the first recorded module (neurosciences) than for the second. The latter being an integrative module in which students would be revising and integrating previously acquired concepts. A possible explanation might be the effect of the content area of the case, or case specificity, here expressed by the module. The neurosciences module requires the use of more technical vocabulary than the integrated module where students might be more concerned about the process of making connections between knowledge. However further research would be required gain a better understanding of this.

For year one the use of technical language increases as the students progress through the different clinical cases. This increase is also visible as students' progress through the sessions of the second module, while in the first module it seems to be the common words that increase. No similar pattern was found in the second-years, these students seem to make a more content use of technical and common language through time.

One alternative explanation, for these observations might be that the second-year group has become more familiarised with the PBL process and is now able to focus more on the knowledge to be learned than the process of exploring the case.

PBL is as much about the content as it is about the process (D. Wood 2003). Second-year students have been more exposed to the process and therefore are more comfortable with it and more able, at this stage, to focus on the acquisition of biomedical knowledge (University of Nottingham 2011a).

First-years are clearly more focused on the case (as the content analysis results show), history and patient and the second-years on basic sciences knowledge, diagnosis and treatment. This would be supported by the differences found in categories of technical terms with second-year using significantly more words related with Diagnosis (for a $p<0.001$), Surgical, treatment and chemicals (all statistically significant for $P<0.0001$). Additionally. Nevertheless, their technical language is

developing as they progress through the cases, as seem by the positive correlation between the clinical cases and technical language, as knowledge is being acquired and used in the discussions.

A more elaborated structuration of knowledge networks by function of the exposure to clinical cases in line with an early or pre-encapsulation stage, seems to be a possible explanation for our results. It would explain the increased use of technical language in the discussions and since language is denser in meaning with few words now being enough to explain complex processes that before would require the use of more non-jargon, the relative frequency of the words per hour (density) decreases.

Aim 2: Understand the strategies used by students with little medical knowledge to discuss clinical cases.

According to the proposed structure of the PBL cases, it was expectable to find a high incidence of questions mainly on second sessions. These sessions are devoted to clinical data enquiry, exploring and questioning of clinical data relating this new information to previous information collected on the case in session one. Our data supports this expectation showing that the highest frequency of questions in both years occurs in the second session. Similarly, there are visible differences in the averages per case between year one and year two. Year one students are using a constant percentage of questions all the cases, while in year two discussions there are visible differences, ranging from 1.52 to 4.61. Furthermore, in general, year two students use a higher percentage of questions. However, for both the different sessions, cases and years differences found fail to reach significance. One possible reason for this maybe the different number of sessions analysed per years as less second-year sessions were captured and analysed, leading to a skewed the data in favour of the first-year as more data is available for calculation.

An interesting finding is that overall first-year students are engaging significantly more in reasoning and explanations then year two students.

Some possible explanations maybe suggested for these findings. First these maybe a product of different group dynamics and individual differences, such as personality and disposition for verbalisation allied to differences in engagement in the process. Although from the viewing of the videos, the conversations with the groups, and interest shown in the study, no reasons where found to believe that major differences in terms of engagement existed, certainly both groups seemed to be

eager to learn and committed to the process. Nevertheless, less engagement of year two could explain the lower density of words in year two, because the students are less willing to discuss the cases. However this group also has a slightly higher frequency of questions, and shows an increased use of technical vocabulary.

These results can perhaps be better explained by an adoption of a more focused and strategic approach to the cases based on their increased familiarisation with the PBL process, as hypothesised in aim 1. A central element to PBL is the opportunity for students to self-direct their learning, making the most of the conditions created by the school and the facilitators (Wood 2003; Barrows 1996). The autonomy of the groups regarding their learning objectives makes PBL discussions very diverse, even within the same institution. Research into PBL should be aware of this fact (Koschmann & MacWhinney 2001). Second-year students have already being exposed to PBL cases and to the process, it is expected that, as they are exposed to the sessions and cases they learn how to maximise their learning opportunities, therefore adopting a strategy of exploring the cases that would suit them as a group best. Maybe first-year students were still learning how to manage the process to best suit their needs and that would justify why there is so little variation of the frequency of questions among cases in year one.

Finally other possible explanation driven from the previous one is that, second-years are not only being more strategic but are so because they are relying more frequently on familiar features of previous PBL cases to approach new cases. As Young and colleagues suggest, even novices rely on familiar features of cases, that are not necessarily the whole case or diagnosis, to approach and make decisions about new cases (Young *et al.,* 2011). Second-year students have been exposed to a pool of clinical cases, they can be relying, more frequently then year ones, on familiar features of those cases to approach new cases. This allows the discussion to be more focused on the critical aspects of the case, the discriminate aspects that can help differentiate between diagnoses. Hence, the less number of words per hour, the higher density of technical language and as the ECA as showing more technical related 'themes' in the discussion. According to this view year one students would be less likely to do this, as therefore they would have to engage more in discussion of the case materials (e.g. history was one the ECA 'themes') until the critical elements of the case are identified. This would explain the higher frequency of explanations and reasoning in year one.

In summary, our results seem to indicate that there is an early encapsulation that might be occurring in PBL students, maybe due to exposure to clinical cases. Additionally these results indicate a differential approach to the cases used by first- and second-year students, but no differences regarding the frequency of their reasoning. First-year students are sharing/using more explanations and second-years slightly more questioning. We present the hypothesis that this might be caused by the extensive exposure to clinical cases during PBL scenarios, increasing students ability to focus the discussions more on differentiating between key features of the case (e.g. particular results in clinical investigations) and diagnoses for the case, which is an approach typically linked with higher diagnostic accuracy (Papa et al. 1990; Eva *et al.* 2002).

Another important finding concerns the method used. We are now confident that corpus analysis is able to be applied to medical education discourse, being useful to "open the black box of PBL" (Hak & Maguire 2000) increasing our understanding of its effects and contributing to research in both linguistics and medical education fields.

It is important to highlight that our corpus analysis results are not aimed at being generalisable. A much larger corpus, a different number of PBL medical schools, countries and many more clinical cases would be required if one wished to extract generalisable results. However, these results provided a useful insight into the way early on in their curriculum students approach and solve clinical cases. This is to be further discussed along with the results of the study two of the present research in the final discussion of this thesis (Chapter 8).

Chapter 6: Comparative Study (Study 2) – Methodology

"Theory, practice and research methods constantly interact to provide a dynamic interplay that helps move theory and practice forward" (Bordage, 2003 p. 1117)

Chapter Summary

The first part of this chapter presents a description of the purpose and methodology of a cross-sectional quantitative study conducted as part of the present PhD research (study 2), aimed at comparing the diagnostic reasoning ability of two groups of students, one at the beginning of their clinical phase and other in their final-year, in the three medical schools described in chapter 3.

The second part will describe the development and validation of the clinical reasoning test (CRT), a cased-based, theory-driven instrument designed for the purposes of this research. This section will begin by presenting the need for this instrument based on the available literature and the steps taken to develop two CRT cases. CRT case 1 has 28 questions and CRT case 2 has 26. Each CRT case had 8 short answers, 14 multiple-choice questions (case 2 has only 13), 2 diagnostic matrices, 2 Likert scales and 2 script concordance type questions (case 2 has only one). These questions were divided into three categories: case questions, research questions and feedback questions.

The instrument will then be presented and its main characteristics and distinctive features outlined and the steps taken to validate it described and results discussed.

A. Methodology

A.1 Philosophical approach

This study is in line with what was outlined in chapter 1 and finds the justification for its methodological choices grounded in the epistemological and ontological views of the pragmatism. This approach fundaments our methodological choices of combining a mixed-methods approach (study 1) with a quantitative study (study 2) now presented in order to gain further understanding of the process involved in the development of clinical reasoning in medical students at undergraduate level.

The study described here is a quantitative, cross-sectional comparison of two groups of students at three different medical schools based on their results on a diagnostic reasoning test, developed and validated for the purposes of this research (CRT). The principles for quality and rigour observed in this study reflect both the assumptions of the pragmatism philosophical view, study design and methods were chose by the ability to provide meaningful answers to the research questions, and the ones defined for qualitative research, mainly: validity, reliability and objectivity. Those and the ethical considerations of the present study will be described in detail in the appropriate sections below.

A.2 Study Aims and Research questions

The purpose of the current study is to understand the impact of educational factors, at the curricula level, in the development of diagnostic reasoning ability in medical students in undergraduate medical education. Our aim is to investigate possible differences in the students ability to deal effectively with clinical cases and strategies used to select information, generate hypotheses, consider possible diagnoses and select between those based on their likelihood for the particular case.

Research questions drawn from the literature review presented in the background chapter guided the research design of this study. These questions are presented below and will then guide the discussion and conclusions of this study.

1. What is the impact of the three types of medical curricula on students' diagnostic reasoning strategies at the beginning and end of their clinical phases?

2. Are there any differences in students' clinical reasoning ability as measured by the CRT?

3. What is the impact of clinical practice/exposure in these differences and on clinical reasoning development?

4. Is the students' knowledge of disease mechanisms (pathophysiology) correlated with students' ability to diagnosis that disease?

5. Is there any correlation between students' results and their confidence in their performance? Are the better students more confident?

6. Do students feel that the CRT is a useful tool to assess their clinical/diagnostic reasoning?

A.3 Study Design

The present study is a **cross-sectional multi-site comparative study between three different medical schools** (Derby, Nottingham and Coimbra) to assess their students' clinical reasoning and compare the clinical/diagnostic reasoning performance in the CRT (presented in second section of the present chapter).

After validation the final CRT used for this research is composed of two parts, each representing a clinical case based on the prevalence of diseases in Portugal and the United Kingdom, guidelines for Undergraduate Medical Curricula in both countries and their distinct difficulty level. A more common case, myocardial infarction and a slightly less common case, diabetic ketoacidosis, were the cases used. The instruments were first designed in English, then translated and evaluated by experts in Portugal, then back-translated into English, and then checked that the two versions of the instrument were identical. This process was repeated during the validation study every time a change was made to any of its versions in Portuguese or English. This ensured that both groups of students would answer the CRT in their native language eliminating possible language bias.

The CRT case 1 has 28 questions and the case 2 has 26, 8 short answers, 14 multiple choice questions (case 2 has only 13), 2 diagnosis matrices, 2 Likert scales and 2 script concordance type (case 2 has only one). These questions are divided in three categories: case questions, research questions and feedback questions.

Case based questions (15 questions in case 1 and 13 in case 2) are the core of the CRT. These are the questions following the clinical cycle from patient presentation to

diagnosis. The purpose is to assess the student's ability to use clinical reasoning to solve a clinical case, by selecting information, developing hypotheses, testing, elaborating and reformulating these hypotheses.

The research questions (10 questions) were designed to allow the identification of the reasoning strategy the students are using, by providing additional information about the process of answering the cases, such as information needed to make diagnosis and moments of decision.

Finally the feedback (3 questions) questions provide us with feedback on whether the student feels that the instrument is assessing what it was designed to access (face validity), whether they value this type of instrument as preparation to deal with clinical cases in clinical environments and if they would like to have access to more cases in this format.

A.4 Sampling

The theoretical population of the current study can be considered as all undergraduate medical students in one of the three most common curricula in medical education: Traditional, Integrated by Systems, and Problem-Based learning curriculum.

The accessible population are medical students in three medical schools.

The sampling frame was defined by the exposure to clinical practice, therefore two groups were selected, one with limited exposure and one close to graduation. This decision was motivated by the evidence provided in literature of the impact of clinical practice (Atkinson et al. 2011; Guenter *et al.,* 2011; Ajjawi & Higgs 2008; Norman 2005) in the development of clinical reasoning (Explained in detail in chapter 2), selecting these two groups as our framing sample would allow to understand if there was, or not, a differential impact of the exposure to real cases based on the curriculum model adopted by the schools. All the students within this frame were invited to take part in the research as volunteers, that is considered a non-probabilistic convenient sampling method.

A.5 Ethical considerations:

This study received approval from the ethics committees from all three institutions involved[37].

Volunteer participation

Previous to volunteering all participants received an information sheet for health volunteers explaining the details of the study, its aims and how the data would be used, stored and managed during the process. Advertising and data collection moments were designed to avoid any type of pressure and ensure the decision to participate in the study was done on an individual and informed basis. Announcements were made on several different occasions and using a range of communication tools, via posters in the different medical schools, in lectures, by email and using available learning portals in the different schools. In the Portuguese medical school, the collaboration of year representatives was asked for posting the announcements on managed online forums. The year representatives' generous collaboration was crucial to ensure that the information about the study reached all the students in the selected cohorts. These online forums are used by the year representatives to post announcements about lectures and other relevant information regarding their studies, therefore they are an excellent way of communication with the students.

A summary of the information about the study, clearly stating the research purpose, names of the researcher and the supervisor and confidentiality of the data was added to the first screen of the CRT. After this information was given and before any questions were asked participants volunteering in the study had to give their consent. This was mandatory, it was not possible to enter the CRT without answering to it, to avoid any nonreplies. Although it was possible to answer the CRT if the 'no consent' option was selected, any answers from these students would be removed separated and removed before exporting the data from the Rogo. The option of allowing students who did not consent for their data to be used was available to ensure that all students could have access to the CRT in fair and equal grounds if they

[37] University of Nottingham, Medical School Ethics Committee (Reference number U/09/2010) and Ethics community of Faculty of Medicine of University of Coimbra.

so whished, independently if they were or were not in agreement with their data being used for research purposes.

Finally a question was asked at the beginning of the test regarding volunteers' previous participation in any of CRT validation phases, this, along with the students email, was used to access the CRT allowing the researcher to identify and eliminate any answers from those who had previously taken the test from the final data analysis.

Confidentiality and data protection

Data collected was kept in a confidential manner. Data exported from the software was identifiable through students' user names and system IDs, for these reasons special care had to be taken, to ensure confidentiality of all results. First, the raw data was not shared and was only accessible to the researchers and her supervisor for audit purposes if necessary, and it was kept in password-protected files in the university servers. Also feedback and results of the CRT results were only communicated to the individual participants, if he/she wished to, and not shared or made available to anyone else involved in the process of teaching or assessing the students in any of the schools.

The anonymisation process prior to data analyses (described below) ensured that results could be discussed and shared among those advising this research without breaching the confidentiality and data protection.

Anonymous data reporting and analysis

The students could ask for feedback on their results, the data collected was not anonymous, rather the anonymisation was carried out after data collection and prior to any analysis or marking of the results.

Raw data exported from the software was prepared, firstly by eliminating any replies from students who denied consent for the study participation, and secondly by replacing all students identifiable information (that is their system IDs, students numbers and emails) with a random research ID. A password-protected spreadsheet containing the match between the research IDs and the identifiable students information (student numbers and email addresses), 'a key spread-sheet', was kept in an external hard-drive in a locked location accessible only to the researcher and the supervisor. All the marking, data analysis and reporting were made on anonymised data, using research IDs.

As students were given the option of receiving feedback on their results, for those who selected this option, and only those, once all the marking was done, the researcher used the 'key spread-sheet', to replace the researcher IDs with the students' emails and sent them their results and feedback. This was generated in a new spread-sheet, containing students' emails and marks, that was stored under the same conditions as the 'key spread-sheet', again being only available to the researcher and for audit purposes the supervisor.

A final ethical principle observed in the current research was the ethical conduct of the researcher, by adhering to the above principles and conducting the analysis, reporting and discussion in a rigorous and true manner, and ensuring that during all the research participants were treated with respect and that their bounteousness in taking part in the present study could not in any way be detrimental to their current or future learning experience.

A.6 Data collection Procedures

Recruitment of students proved to be a very difficult task. Students in all schools have very demanding curricula in terms of time and workload and limited availability. Additionally, the fact that the CRT would take approximately one hour (for both cases), may have contributed towards these difficulties. Nevertheless, all efforts possible were made to try to increase participation.

All the students in the defined cohorts (BMedSci/BMBS 5th semester/final, GEM 4th semester/final, FMUC 3rd years/final) were invited to take part in the study, by email, by announcements made in lectures, by announcements in intranet and posters placed in strategic places in medical schools. Participation was incentivised by offering a symbolic prize for the best (Amazon/Fnac vouchers to the value of £25) results in each of the cohorts, individualised feedback on their performance and model answers from experts.

Additionally, to increase participation of the CP1 cohort in the UK (both GEM and BMedSci) medical schools, the CRT cases were made available as a progress test. Progress tests are a common practice in these schools; students are not obliged to take them, neither are their results used for summative purposes. These tests are an additional formative resource the schools provide to the students to help with their study and revision for exams. It was agreed that both cases of the CRT would fit this purpose and therefore permission was given to upload CRT as two separated cases,

case 1 and case 2, as progress tests on to the students NLE (Network learning environment). An information sheet clearly explaining the details of the study and a consent form were also uploaded to ensure the students knew and consented to take part in the study. Students who agreed that their data could be used for the research would then have to select the appropriate option, if they did not agree the CRT case would still be available to them to answer but their responses would not be included in the response database. This was not initially planned and it was a measure to try to improve recruitment numbers. Indeed this proved to be one of the most effective strategies to recruit students from this cohort. However, the CRT cases had to be uploaded as two separate progress tests due to their length. Although some efforts were done to contact the students and incentivise their participation in both cases, it was not possible to guarantee that the same students answered both cases.

In the Portuguese medical school progress tests are not common practice in this school, neither is online assessment and there is no network-learning environment in place. Therefore, as an additional strategy, some key lecturers in the medical course were identified where information about an email link to the Portuguese version of the CRT was presented to the students. The same additional procedure was repeated for the final-year Portuguese students. Finally, the researcher directly contacted some FMUC student representatives asking them to publicise the study among colleagues. Due to technical problems it was not possible to use this system with the Portuguese students, however, similar software, Surveymonkey, was used with these students. This system was chosen for its ability to create a very similar in all the functionalities, but also, allow the researcher to edit aspects such as, look, fonts and colours in order to make it as similar as possible to the Rogo environment. Data from Nottingham and Derby students were collected using Rogo, an online assessment system developed by Dr Simon Wilkinson in the Medical Education Unit of University of Nottingham. Students' answers raw data was identified by the student number, to detect any repetition and to ensure data from students who took part in the pilot phase was not included in the final sample, as these students had previous contact with the case and its answers. After initial checks, results were anonymised (further details in the ethical considerations section below).

Table 6. 1: Dates of data collection

Cohorts		PBL & BMedSci	FMUC
CP1	Case 1	04.04.11 to 15.04.11	31.05.11 to 29.06.11 (both cases available)
	Case 2	10.06.11 to 24.06.11	
CP3	Both cases	01.06.11 to 24.06.11	31.05.11 to 29.06.11

Exclusion of students who had taken part in previous validation studies (described in the second section of this chapter) was made using the students automatically generated system ID and their student numbers. Additionally, a question was placed at the beginning of the CRT, asking students about their previous participation. These students could still take the CRT and receive feedback, however, their responses were not included in our dataset for the research purposes.

A.7 Data analysis Procedures

Answers collected electronically were exported to *.cvs* files and Microsoft Excel format, anonymised for research purposes by removing any data that could lead to the student's identification, as aforementioned in the ethical considerations, and then marked individually by the researcher using a set of defined objective criteria and model answers previously developed. Student identification was kept in a separate file, allowing a match between the students research ID and their personal information once all results were marked.

A first analysis of the quality of the exporting process and the integrity of data was carried out. Data were organised according to the questions in the CRT, marked by the researchers using the marking criteria agreed with the experts and the results were copied to a separate database. Final marks to each of the questions were uploaded into SPSS (statistical software packed for social sciences) 19.0 and analysed. Three separate datasets were created, one with the answers to the questions presented in case 1, one with all the answers presented in case 2, a third one with the answer to case 1 and 2, plus the answers to the research questions and feedback. The data inspection was carried out, followed by missing value analysis and replacement. Datasets were organised by year and courses and missing values were replaced by the *mean of nearby points*. Analysis of variance (ANOVA) was carried out between gender, year and course (medical school) for case 1. Case 2 results failed the

normality tests, therefore non-parametric statistics were used, mainly the Kruskal-Wallis Test. Significance was defined for $p<0.05$. Correlations between results in both cases were performed using Pearson's correlation. Additional correlations with research questions were also performed.

A.8 Criteria to ensure rigour and quality

The criteria for rigour adopted are aligned with those suggested in the literature for quantitative studies, mainly validity, reliability and objectivity (Golafshani 2003).

Objectivity, validity and reliability should also be applied to the research or the instruments/tools used as valid and reliable research always requires the use of valid and reliable instruments (Neuman 2003).

Objectivity, within the positivist or post-positivist paradigms is relates with the replicability of the research (Salkind 2010; Neuman 2003). In educational studies, replicability is not possible, neither desirable in most of the cases, as the contextual factors are of extreme importance to our research and to the impact of its findings (Berliner 2002; Creswell 2009). As stated in the introduction, the current study is situated within the pragmatic approach, not the pure positivist paradigm. Objectivity here is defined as ensuring the alignment between the research questions and the methods used, by ensuring enough justification for the methodological choices and well explained arguments to support any interpretations made considering the context in which this research took place, as well as its purposes (Creswell & Clark 2007; Biesta & Burbules 2003; Morgan 2007; Feilzer 2009).

Validity is concerned with the extent research is investigating what it was designed to investigate, meaning research methodology (instruments used, data collected, analysis performed) must be coherent with the research questions posed (Salkind 2010).

The main types of **validity of instruments or tools** can be defined as (Neuman 2003; Salkind 2010). ***Face validity:*** This deals with participants' opinions and perceptions regarding the instruments ability to measure what is supposed to be measured (Tavakol & Reg Dennick 2011). ***Construct validity:*** This refers to the relationship between the scale or scores in a test and the theoretical basis of the construct being measured. ***Convergent validity***: Is related closely with construct validity, and is the extent to which an instrument fits the nomological networks of related constructs described by theory. For example, it is known that memory test performance

declines with age, therefore if studying an instrument known to be directly correlated with performance in this test, a similar relationship with age would be expected, and this would mean the instrument has convergent validity. ***Concurrent validity:*** Represents the correlation between the results of a test/scale and results obtained using other previously validate measurement for the same construct. ***Content validity:*** Is related to construct validity and in simple terms refers to the ability of a test or scale to measure a certain construct in all its features or aspects. This is often verified using experts in the field.

In this research the validity of the current study was addressed during the validation study, described in the section below. Here several modalities were used: focus group, expert review and pilot data, to ensure the instrument created had sufficient validity. All these procedures were previous to final data collection and aimed to ensure the quality of the data collected.

Reliability is a measure of the internal consistency of an instrument or scale. The most common measure is Cronbach's Alpha statistic, with an optional associated factor analysis (Tavakol & Reg Dennick 2011). This was developed by Lee Cronbach in 1951, and calculates the internal consistency of a test by investigating the relationship between internal scores (intra and inter-cases) and between those and the total score of the test. This is expressed on a 0 to 1 scale with the generically accepted value of 0.6 to 0.7 being considered as a minimal value and values over 0.8 considered as very good values for good internal consistency. Factor analysis tests the homogeneity of the test, which contributes towards its reliability. By analysing the scores on the different items in a test, and the relationships between them and the final scores, factor analysis identifies items that cluster together showing how much each item in a test is contributing to the final result (Tavakol & Reg Dennick 2011).

Reliability measures were calculated using Cronbach's alpha. As above these procedures were previous to the application of the instrument, and ensured that there was sufficient evidence that the CRT could gather the necessary data to answer the research questions posed.

B. CRT Development and Validation

B.1 Need for a new instrument

Clinical reasoning is a highly complex process (Croskerry 2009a) and as with all mental processes it is not available for direct observation or measurement increasing the difficulty of its assessment (Holyoack & Morrison 2005). Previously in the background chapter we described the instruments and methods used to assess clinical reasoning. These were divided into instruments for experimental/research purposes and those designed for the purpose of assessment. Experimental methods, such as thinking aloud protocols or protocol analysis, have been extensively used in clinical reasoning and mental process research, as they provide an in-depth view of the participants' thought processes (Patel, 2005). However, these methods are not suitable for use with large groups of participants (Ericsson, 2006) as the level of analysis required is too expensive and time-consuming. Furthermore, these methods are likely to be influenced by cultural and linguistic aspects (Ericsson, 1993). As our aim is to compare clinical reasoning outcomes from students in three different medical schools, in two different countries in a transcultural comparative study, our research required a method or instrument that would be less susceptible to language differences and that could be used with large samples. Therefore, although considered initially, such methods as thinking aloud or protocol analysis were considered not to be the most appropriate for our purpose.

Several instruments have been designed to assess clinical reasoning during the last 40 years of research in this field, from patient management problems (Rimoldi, 1955), including tests of diagnostic skills (McGuire, 1985), simulated patient-problems (Helfer, 1971) and Programmed Tests (Hubbard, 1965) to more recently Script Concordance Test (SCT) (Charlin *et al.* 2000) and CIP (comprehensive integrative puzzle) (Ber 2003) and many others previously been described in the Background chapter. Although advantages can be identified in each of these instruments none seems to provide a valid tool for the purposes of this study, that is, allowing comparison between undergraduate students at different levels and in different curricula. Furthermore, none are able to reveal differences (if they exist) in overall performance and simultaneously provide information about potential differences in the reasoning strategies used. For these reasons we developed the Clinical Reasoning Test (CRT) based on the main strengths of some of the above instruments,

recombining some of their features and/or principles in order to minimise the weaknesses identified by research. The next section will describe the steps taken in order to design the CRT.

B.2 Clinical Reasoning Test (CRT) Development

Developing the CRT was an iterative process before the final version used in the pilot study (presented in the next validation section). Eight versions were developed in constant consultation with a panel of experts from all the schools taking part in this study.

The CRT was designed to be an online test, since its early stage questions started to be tested and content started to be uploaded to the University of Nottingham's online testing system Rogo,[38] developed by Dr Simon Wilkinson. This allowed us to focus simultaneously on the usability and user experience perspective of the CRT while developing its main features, ensuring both processes were aligned.

B.2.1. CRT Structure

A first step in the design of the CRT was to perform a critical analysis of existing instruments, identifying their main strengths, weaknesses and usefulness to achieve the aims of the present research. Based on the literature dealing with the principles of clinical reasoning assessment and previously developed instruments some aspects were identified.

The Diagnostic Thinking Inventory (DTI) marking scheme is a very useful and widely used tool, however, it is associated with a self-reported instrument. In the CRT design process our aim was to design questions that could provide evidence for each of the aspects considered in the marking scale of the DTI but replacing the self-reported aspect by a more objective measure of performance. The Key features (Bordage, 1990, 1995) approach was used in order to identify the key aspects to be identified in the scenarios/cases presentations and to design some of the questions, mainly

38. Rogo: https://rogo.nottingham.ac.uk/

related to history questions and clinical investigations. Clinical reasoning problems (CRP) (Groves, 2002) clearly emphasise the importance of making predictions related to clinical data and the importance of being able to distinguish the ones supporting or opposing a diagnostic hypothesis. Therefore a question using this scheme was introduced into our template.

Finally some script concordance test (SCT) (Charlin, *et al.,*2000) type of questions were introduced in order to assess students ability to deal with new information, mainly related to clinical investigation results.

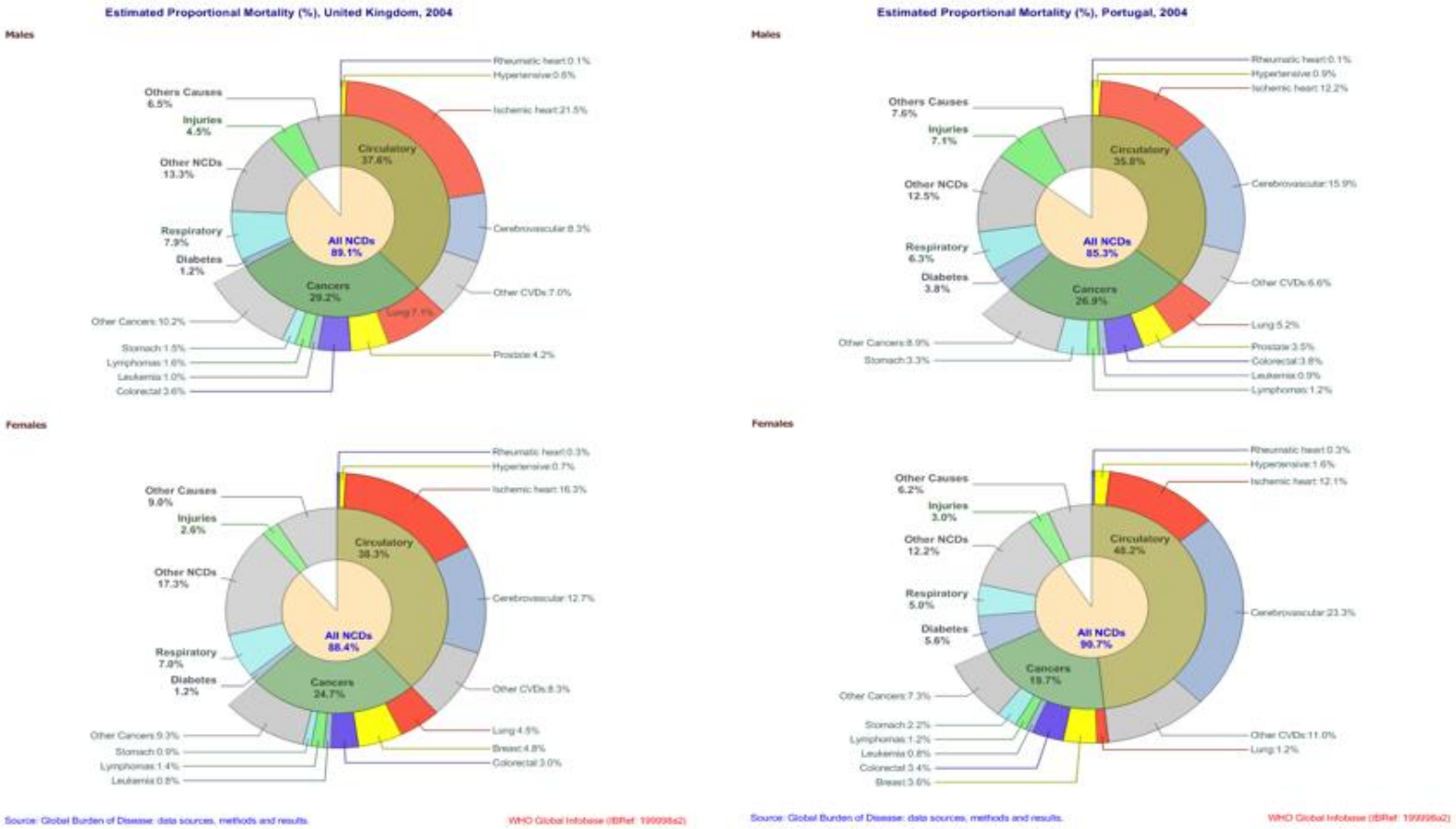
Additionally, at this stage, the choice of using one long case with multiple questions or several multiple case presentations with only a few questions each was considered and a choice to use long cases was made. Both options had advantages and disadvantages, however, our choice was made on the basis of the one that best suited the aims of this research, namely not only comparing clinical reasoning performance but also identifying differences in the strategies used by students in order to achieve that performance. Longer cases, that cover all the aspects of the clinical cycle, from initial presentation to diagnosis, that allow one to gather data on students' diagnostic performance as well as on how they perform throughout all the phases that lead to the diagnosis are better suited to answer the research questions posed by this study.

Additionally some research questions were added to the CRT structure, in order to collect data regarding some additional aspects that were not being covered, and were considered important for our research (e.g. confidence about the diagnosis).

Selection of clinical cases

After developing a draft structure for the CRT a few cases were selected on the basis of their prevalence for both UK and Portuguese populations based on information from the World Health Organisation (Mathers & Fatt, 2008) and the medical curriculum outcomes in both countries as defined by their regulatory bodies (GMC, 2009). Four cases were identified: a myocardial infarction, a diabetes complication (Diabetic ketoacidosis), a chronic respiratory disease (Chronic obstructive pulmonary disease), and a gastroenteritis by *campylobacter*.

Figure 6. 1: Estimated proportion of mortality in Portugal and the UK (2004) extracted from (Mathers *et al.,* 2008)



Key individuals as listed earlier in all three schools were consulted in order to confirm that those cases would be aligned with curricula delivered and their difficulty levels would be appropriate for their students. This decision was motivated by feasibility reasons, as it was not possible to blueprint the clinical reasoning test against the curriculum. The level of detail and type of curriculum map used in the different schools would not make this a productive task. Asking those involved in teaching the students proved to be a much more useful way to gain insight into the adequacy of the cases and the test.

During the validation process it was shown that answering all four CRT cases was too time consuming for the students, and that two cases had weak reliability, therefore only two cases were included in the final version of the CRT used in the present study: myocardial infarction and diabetes complications (Diabetic ketoacidosis).

Combining structure and content

The next step in creating the CRTs was to introduce the content into the previously defined structure (see Appendix 3: Example of guide for CRT cases). In order to do so several sources of information were used: the NICE guidelines (NICE, 2003; NICE, 2010), NHS choices Map of Medicine, which is a comprehensive evidence-based map of the points of care and medical decisions regarding common and important conditions (NHS, 2010), Clinical Knowledge Summaries[39], which are summaries of evidence regarding best-practices in diagnosis and treatment of diseases provided by the NHS for medical professionals; and selected research papers on the considered themes. These sources, although UK based, are used by professionals world-wide including in Portugal, therefore were used to construct the draft version of the cases, along with, when necessary, searches in Portuguese health ministry website. Experts in all medical schools were consulted in order to verify the clinical content accuracy of the questions[40].

Translation and back-translation

Until this moment the development of the CRT cases was done only in English to avoid confusion between versions. This was possible because the advisors of this project in Portugal were bilingual and proficient in English. At this stage a translation

---

39. NHS, Clinical Knowledge Summaries. Available at: http://www.cks.nhs.uk/home.

[40] Details about this process will be provided in the section review by experts.

of the CRT cases was carried out by the researcher, and a Portuguese version of the cases was produced. The Portuguese advisors were asked to comment on the translation and to identify possible aspects that would not be suitable or realistic in the Portuguese context. After this the Portuguese version was back-translated into English and compared with the original version in order to test the accuracy of the translation. This process was repeated for every new English or Portuguese version of the CRT cases.

Cultural differences in practice: Minor adjustments

During the development stages, some aspects were identified by the Portuguese advisors as not being clear for the Portuguese context. One was the format of the clinical investigations as their organisation (tests grouped together and procedures included in each test) was slightly different in the Portuguese context. Therefore the table of clinical investigations had to be changed to a neutral and simplified format. The style of some questions was thought to be confusing to students in the Portuguese version, therefore the terminology was changed. For example the questions *"What would be your main concerns about this patient?"* was replaced by *"List the main clinical problems".*

Pilot stage

A pilot version was developed containing four cases and delivered to the students in three sessions arranged in order to test the psychometric characteristics of the instrument, but also its usability and to gather feedback. Also this was an opportunity to test the data collection procedures and data exporting methods. A table with data collection dates, times and participants will be presented in the results of this section.

Final version

After data from the pilot was analysed some alterations were made and the final version of the CRT cases were reviewed by the expert panel to ensure their agreement before data collection.

B.2.2 Clinical Reasoning Test Main Features

Theory-driven

As described before the CRT is based on the literature concerning other clinical reasoning assessment instruments, mainly the Diagnostic Thinking Inventory (Bondage, 1990), Key Features approach (Bordage, Grant & Marsden, 1990). Clinical Reasoning Problems (Goves, 2002) and the Script Concordance Test (Charlin, *et al.,* 2000b). The CRT was designed on research conclusions regarding each of these assessments simultaneously minimising aspects that had been criticised, or proved not to be valid, and applying what had been highlighted as valuable to clinical reasoning assessment in these instruments. This makes the CRT distinct from previous clinical reasoning assessment instruments, as these made limited use of previous instruments in designing new ones. Also, the CRT differs substantially from simple clinical cases that were not specifically designed for the purpose of clinical reasoning assessment according to the literature of this subject. The theoretical basis of the CRT also provides a useful framework to interpret and discuss the results of the validation process.

The combined use of different clinical reasoning tests, was also used in a recent study by Amini (2011) in Iran, where the authors used four different clinical reasoning tests (Key features, Script concordance, Clinical reasoning problems and comprehensive integrative puzzles) as a multi-instrument assessment tool of top-ranked medical students in the Medical Sciences Olympiad. The authors concluded that the combined use of different clinical reasoning instruments is a reliable way to assess clinical reasoning (combined reliability 0.91) (Amini, 2011). Although some aspects require further analysis. First the feasibility of using this combination out of the particular context of the medical sciences Olympiad seems to be very limited. This assessment lasted a total of 16 hours during a two days period to be completed and involved a group of 15 experts along with a research team with 20 members. Also it is not clear how all these instruments were combined, or if each item addressed a different clinical case or different parts of the same case were used in the different formats. Furthermore the results form the test battery fail to reach a high correlation with the students' university grade point average, which casts doubts about its validity, especially because no other measure of validity was included. Nevertheless,

this study shows that it is possible combining different instruments to assess clinical reasoning without compromising reliability of the assessment.

Long cases

The CRT is based on long cases in comparison to short cases where only one of two questions are asked regarding patient presentations or scenarios. This decision was motivated by the aim to increase realism of the assessment and because long cases were deemed to better suit the purposes of this research. Although recognising that the results of these two cases cannot be interpreted as predictors of student performance on future cases due to case specificity (Elstein, 1999), long cases allow for more insight into how a diagnosis is achieved by analysing previous decisions (Geoff Norman 2002b). Additionally, because our study is a cross-sectional comparative study, our main aim is to design an objective and valid measure of clinical reasoning outcomes that allows comparison of students in three different groups while simultaneously identifying differences in the strategies use to achieve those outcomes. Finally, long cases provide an integration of all the parts of the clinical cycle improving the authenticity of the testing (Norman, Patel & Schmidt, 1990).

In the CRT final version[41] there are 15 (case 1) to 13 (for case 2) case based questions that start from the patient presentation and cover the steps of the clinical cycle from presentation to diagnosis.

Simulated patient presentations (audio recorded)

The initial case presentation is made through an audio recording of a simulated patients' initial complaint. This feature was designed with the purpose of testing the student's ability to identify important information from a "real" case presentation and not a summarised version where only the important parts are mentioned. This second type, often used in case-based scenarios and examples of clinical cases, provides the students with an already processed summary of information where the main key concepts can be easily recognised by the students (Eva, 2005). Also summarised vignettes tend to make use of medical terminology and not the lay terms in which a patient's presentation happens in "real life" (Eva, 2005).

41. The previous versions of the CRT included more questions that were eliminated due to the validation process, described in section E. of the present chapter.

Additionally audio files add realism to the case, adding additional clues that should be identified, processed and used by the students, as they are by experts, to then guide them in the history taking process (Eva, 2010).

Different question types:

"Question types should be selected according to their specific strengths and weaknesses. A good assessment procedure therefore consists of various methods, each of which should be tailored to the specific needs of the assessment." (Schuwirth, & van der Vleuten, 2004 p. 977).

The CRT cases incorporate different types of questions: 14 (13 for case 2) multiple-choice (some of which yes/no/maybe/or don't know), 8 short-answer, 2 Likert scales, 2 script concordance and 2 (only 1 on case 2) matrices. There are a few reasons for these choices. Fairness to all the students involved in the study is important, as assessments in each of the medical schools differ from each other. For examples PBL students are more used to short-answer questions, while a large majority of the assessments in the traditional curriculum medical school include multiple choice questions only. Including only one type of question could represent an unfair advantage to the students who are more familiar with that format, and that would be a serious limitation of our study (Muijtjens *et al.,* 2007). Additionally the type of question has been shown to have an impact on the reasoning strategies used (Epstein 2007). Extended match questions fostering hypothetico-deductive reasoning while short answers tend to promote more non-analytic strategies (Heemskerk, *et al.,* 2008). Therefore the combined use of several types of questions should contribute to minimise potential bias towards a particular type/strategy of reasoning. Another reason for the use of multiple question types was the differential purpose for what each question was designed to assess (Schuwirth & van der Vleuten, 1995). Short-answer questions were used to assess aspects such as summarising information, justification or reasons, predictions of results and diagnosis. Each of these aspects are understood to be better assessed through open-ended questions, that avoid leading students to recognise the right answer in a list of distracters instead of recalling and applying their knowledge to provide an answer. On the other hand, when assessing the students' ability to select between many options, multiple-choice questions represent the best choice. Script concordance questions have been shown to be a valid question type to assess the ability to deal with new, unexpected information, mainly related with clinical data interpretation (Schuwirth, & van der

Vleuten 2004), and were therefore used in this context. Finally the differential diagnostic matrices are four simple one word open-ended questions presented in a matrix, where students were asked to judge the seriousness and likelihood of the differential diagnosis. Matrix type questions are an appropriate type of question to measure the central analytic process in complex situations (Kyllonen & Christal, 1990). Likert scale questions were used for research and feedback purposes only.

Finally, it has been argued that multiple question types in a test may represent a challenge to validation procedures, as the use of different types of question may lead to a decrease in the internal consistency of the instrument and hence reduce its reliability. However if the test is well designed via a rigorous review and validation processes this limitation can be minimised (Schuwirth & van der Vleuten, 2004).

Online mode (Rogo) & Unidirectionality issues

The CRT was designed as an online test. This delivery mode, allows one to overcome one of major criticisms concerning the use of long cases, mainly the Patient Management Problem type of assessments. In these assessments the information is made available to the students according to the students previous selection, leading to possible *consequential error*[42], where one incorrect decision at the beginning of the case could lead to an incorrect diagnosis. For example, not "asking" an important question in history taking, means that the student would not be given that information in any subsequent part of the case. Consequential error could lead a student to fail, not by lack of skills, but by lack of information (Schuwirth & van der Vleuten, 2004). Therefore avoiding this type of error is fundamental to ensuring an instrument is measuring clinical reasoning.

A way to avoid this error would be to give all the information about the case to all the participants at the beginning of the test. However, this would make it impossible to assess important aspects of clinical reasoning, such as the ability to choose clinical investigations, or appropriate courses of action, the ability to test hypotheses, or how they would change due to new information.

Software allows us to make information available in a sequential order, while not permitting previous answers to be modified in the face of new information. Rogo allows us to have different screens, where only part of the relevant information, for

---

42. Describes an error that is a direct consequence of a previous mistake in a related task. (van der Vleuten & Newble, 1995).

the moment, is displayed (e.g. in a history question). Once the student moves to the next screen, he/she will not be allowed to return and change previous answers; therefore, information presented can be independent of previous answers.

For example screen 1 presents a list of possible clinical investigations and the student is asked to make a choice. In screen 2 (the following screen) the results from only the correct or relevant investigation can be presented to all the students. At this point the student would realise if their previous choice was correct or not, but would not be able to change their choice. This is referred to as the 'unidirectionality' of a test and avoids consequential error. Additionally, online testing is more flexible, allowing for different types of questions and resources (such as the audio files) to be added to the test. It adds accuracy and speed to the data gathering procedures, that are in this case automatically done by exporting options, avoiding errors in data entry.

Online testing, however, cannot be seen as a simple transposition of paper and pen testing to a computer screen, it requires special procedures to be developed (Dennick, Wilkinson & Purcell, 2009; Granello & Wheaton, 2004). The CRT design followed 10 out of the 12[43] guidelines presented by Granello & Wheaton (2004), although these are guidelines about research in general they provide a very useful framework to design of online assessments. How each of the 12 guidelines was applied to the CRT development is presented in the table below. Following these guidelines provides confidence in the data and the adequate use of online testing in our study.

Table 6. 2: Guidelines for the construction of online tests

Guidelines	**Implementation on the CRT development**
1.Determine the population to be measured. Do all members of the population have equal access to the Internet?	Internet access was provided by all the medical schools involved in the study, as computer lab, and free wifi-access on campus.
2.Determine whether an e-mail or Web-based survey will be used.	The CRT was developed as a Web-based survey in order to allow unidirectional presentation of information in different screens.
3.Develop the layout of the survey and the type of format for the questions,	The type and format of questions were justified in the section above. The layout of

43. The remaining three guidelines are not related to construction of online tests but with data collection and organisation.

and (4.) write the questions.	the survey was defined by the steps of the clinical cycle (patient presentation, history, physical examination, differential diagnosis, clinical investigations, final diagnosis).
5. Keep the layout simple, with easy-to-read fonts and a consistent layout throughout.	Rogo (e-assessment platform) provides a template for the questions formatting and generic layout that was used. However after assessment by participants some of the fonts sizes were enlarged.
6. Be sure to address informed consent issues, including the name and contact information of the researcher.	Consent form was e-mailed to the students along with the link to the CRT. A content information sheet and an ethics consent number were included as the first screen of the CRT. After that a question regarding consent was mandatory in order to proceed to the CRT cases.
7. Determine how data will be entered into the computer	Data was entered into the CRT electronically, and the students were informed they could make use of abbreviations. Students were given previous information about the test and the test questions to ensure they were aware of how to answer the questions appropriately.
8. Practice putting in data.	Tests were made by the researcher to guarantee the test was working adequately.
9. Include "error detection".	The CRT cases would automatically register users (the student number for UK and an IP address for the Portuguese students), however once the data was exported all entries that did not contained any answers to questions, or answers from students who did not consent the participation in the study were eliminated.
10. Pilot the study using a subset of the target population.	The next section of the present chapter will describe this process.

The CRT is a theory-driven instrument which means questions were not developed to fit the case but rather to reveal the cognitive processes the students were engaging in. Therefore the type of questions set are independent of the case and can be

applied to any case, scenario or context required. The difficulty level of the CRT cases is then determined by the content of the case and context (situation and respondents), as shown by the figure 6.2 below. The difficulty of a case is related to previous frequent contact with similar cases (Granello & Wheaton, 2004). This the basis of the non-analytic model of clinical reasoning.

If a case is based on a condition that is highly prevalent[44] or frequent in the context of the respondent, and/or the case presentation used corresponds to a typical patient presentation for that disease, the CRT case would be expected to be easy. If on the other hand the case presents a less prevalent condition or a less typical presentation, the CRT case would be expected to be of increased difficulty. The degree of prevalence and the mode of presentation in a CRT depends on national and international indicators but are also a function of the particular person answering the question. For a specialist, prevalence should be understood from experience within a specialty, while for a general practitioner prevalence would be based on the context of the population where his or her practice is located. Also the CRT cases can reflect the career stage of the test population. Certain cases would be adequate for students at a postgraduate trainee level whereas others might be suitable for undergraduate level. This relationship is depicted schematically in Figure 6. 2..

Focus on diagnosis

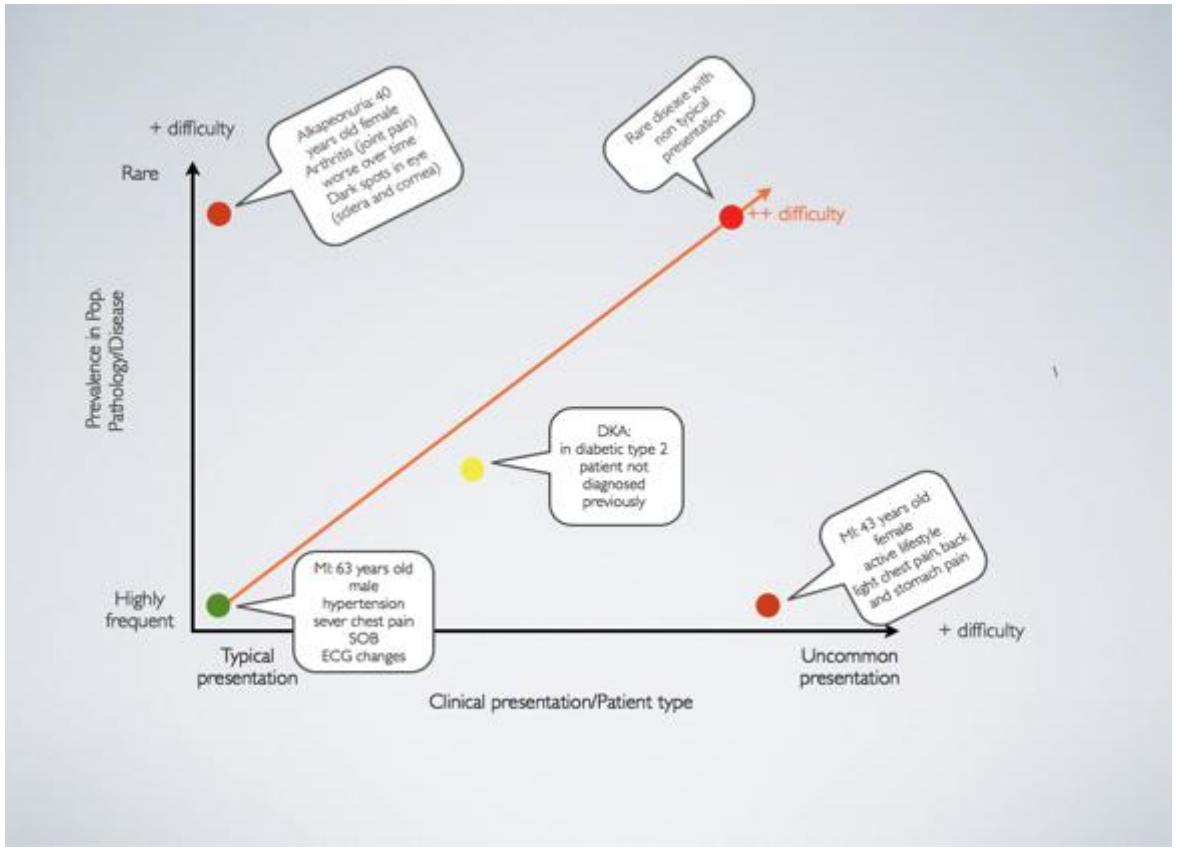
The final version of the CRT cases focuses on diagnostic reasoning. Although the initial CRT cases had questions regarding treatment and prognosis these were removed from later and final versions of the CRT. A few reasons justify this action. First the balance between test length and its aims is important. Data and feedback from the students during the pilot study provided a clear indication that the cases were too long leading to drift in students' attention in the final questions; therefore, some questions needed to be removed. When consulting with experts on their opinion about which of these could be removed without compromising the quality,

---

44. Prevalence must be read here as related to the individual's reality. Example: Tropical diseases are not very frequent in a European population, but for a clinicians working at an Institute of Tropical medicine in the Pacific they might be common, therefore highly prevalent.

realism and the ability to assess important steps of the clinical reasoning process, a unanimous response was that treatment questions could be eliminated. In their opinion, these questions were assessing factual knowledge of pharmaceutical and management procedures. It was also pointed out that, in their practice, when a diagnosis is made the right therapeutic option to follow is often achieved by consulting books or compendia. However, undergraduate students are not expected to be able to deal autonomously with treatment, therefore such questions are necessarily out of the scope of the present research.

Figure 6. 2: Relationship between level of difficulty and prevalence on the CRT cases clinical presentation in the CRT



Sudoku puzzle

The CRT case 1 included a sudoku puzzle as a distractor task prior to engaging with CRT case 2. Distracter tasks, e.g. arithmetical tasks, are common practice in psychology research in order to clearly mark the end of a task while simultaneously allowing time for any information pending in short-time memory to be organised (stored in long-term memory or forgotten) (Nairne, 1992). This is a necessary step when information from a previous task could interfere with subsequent tasks (Nairne, 1992). An example of this interference would be the participants having in mind characteristics of the patient in case 1 while answering case 2. Sudoku puzzles are forms of reasoning tasks, with better performance being associated with better deductive reasoning skills (Louis, Goodwin, & Johnson-Laird, 2008). Therefore its use as a distracter was more relevant to our study then a common arithmetic calculation, as this would ensure that even though it was a distractor task, the students would still be engaging in a reasoning activity. However, due to time restraints students were told not to try to complete it, but just to allow a maximum of 2 to 3 minutes between cases, and for that reason the puzzle results were not marked or used in the present research.

B.2.3. Marking the CRT

Development of the CRT marking scheme was an iterative process with frequent expert revisions and inputs. The aim was to ensure the marking scheme was as objective and fair as possible and that marks attributed to questions were aligned with the aim of the present research. Model answers to the questions were defined on experts' advice and joint thinking/discussion between this group. For the short-answer questions key words in each were identified to reduce complexity and bias of marking (see Appendix 3: Questions/Marks/Model Answers). A marking scheme was developed by attributing marks to each individual question based on its importance to the overall case. Case 1 presentation and features were due to be less challenging for the students, therefore additional (two) questions were introduced in order to make both cases more equitable. Even so, we tried to ensure the balance between marks given to each of the aspects in the clinical reasoning process assessed (selection, hypothesis generation and testing) was similar in both cases (table 6.4).

Hypothesis generation was seen as one of the most important aspects, followed by selection of relevant information and hypothesis testing and data interpretation for both cases.

Objective questions, such as multiple choice questions and script concordance type questions were given a mark on the basis of being the most appropriate answer for this patient scenario at that moment. The total mark for each question was then calculated by summing the marks given to each of the key elements. This was a crucial step in order to guarantee that marking was objective and not influenced by the language skills of the students.

Model answers and key points were defined at graduation level, that being what would be expected from an average final-year student. This was motivated by the need to compare students from different cohorts and schools, therefore a single marking scheme had to be used. By setting standards at the level of an average final-year student it ensured that the test would be challenging for both cohorts, and that differences could be found between cohorts. The first steps taken to define the marking criteria were undertaken in parallel with the definitions of the CRT cases. After the pilot study, alterations were made to the CRT, requiring an extensive review of the marking scheme. This was done by repeating the second step described above, but focusing mainly on the balance between the marks and the aims of the study. At the same time that the model answers were defined the percentage of the total number of marks to attribute to each question was also discussed with experts. The CRT was designed to test clinical reasoning; therefore, the marks attributed to each question should reflect their importance in the overall reasoning process. The marking of the SCT type question was simplified (see table 6.3) this was decided in consultation with the experts to be more appropriated for the stem of the question and the level at which the test was set.

Finally two clinicians, external to the study, were asked to mark a sample of 5 CRT cases each. All markers agreed in 78% of the answers, with only small differences (between 0.5 and 1) in the remaining answers. Table 6. 3 below shows the distribution of marks per area of clinical reasoning as assed by the final CRT cases.

Table 6. 3: Distribution of marks per CRT question (see Appendix 2 for Questions/Marks/Model answers for Case 1 and case 2)

Question	Marks	Type of question	Type of marking
Summarise the patient information you have just heard (limited space 300 words). (Q5)	10 marks 1 mark per key feature	Short-answer	Manual
List here the main clinical problems: (Q7)	3 marks 1 mark per any problems	Short-answer	Manual
If you could choose the course of action to follow, what would you like to do immediately? (Q8)	5 marks correct 4 marks alternative	Multiple choice	Manual
Please explain why would you take that course of action. (Q9)	5 marks for good explanation; 3 marks for vague explanations (e.g. need more information)	Short-answer	Manual
Identify question to ask from the list according to their importance: (Q10)*			
Essential questions (Q10.1)	3 marks 1 mark per correct question selected	Multiple choice	Automatic (software)
Important (but not very discriminatory) (Q10.2)	3 marks 1 mark per correct question selected	Multiple choice	Automatic (software)
Least relevant questions (Q10.3)	3 marks 1 mark per correct question selected	Multiple choice	Automatic (software)
If you would like to ask any other questions to the patient, please write them here: (Q11)	5 marks for any 2 relevant questions containing elements of this case 2.5 vague question ("SOCRATES" was given 1.5 as it is relevant but it is a general mnemonic for a series of questions not a specific question)	Short-answer	Manual
Differential Diagnosis Matrix (Q13)	3 marks for just correct diagnosis 4 marks for 2 plausible answers (not most appropriate column/row) 5 marks for 1 plausible answers (not most appropriate column/row) plus the correct diagnosis 6 marks 1 right answer and a plausible answer (not most appropriate column/row) 7 for 2 right answers, without the correct	Matrix	Manual

Question	Marks	Type of question	Type of marking
	diagnosis 8 marks for 2 right answers, one being the correct diagnosis		
Marks Select from the list 3 results you would expect to find in the physical examination? (Q15)	1 mark per correct selected answer	Multiple choice	Automatic (software)
Differential Diagnosis Matrix 2 (Q17)	3 marks for just correct diagnosis 4 marks for 2 plausible answers (not most appropriate column/row) 5 marks for 1 plausible answers (not most appropriate column/row) plus the correct diagnosis 6 marks 1 right answer and a plausible answer (not most appropriate column/row) 7 for 2 right answers, without the correct diagnosis 8 marks for 2 right answers, one being the correct diagnosis	Matrix	Manual
Chose clinical investigations (Q18)	5 marks 1 mark per correct option selected	Multiple choice	Automatic (software)
Write here 2 results you expect to find in the clinical investigations (Q19)	2 marks per any 2 correct answers	Short-answer	Manual
Script concordance questions (Q20-25)**	6 marks (case 1) 4marks (case2) 1 mark by correct answer (same selection as the experts)	Script concordance	Manual
Based on the entire history, examination and investigations, what do you consider is now the most likely diagnosis? (Q26)	2 marks	Short-answer	Manual
Why do you think that is the most likely diagnosis?(Provide reasons for your answer to the previous question) (Q27)	2 marks per at least 2 valid reasons	Short-answer	Manual
Script concordance questions (Q32)*	1 mark by correct answer (same selection as the experts)	Script concordance	Automatic (software)
Total marks	74 Case 1 / 62 Case 2		

*Questions were only included in CRT Case 1:MI. **Only four options were presented in CRT Case 2:DKA

Table 6. 4. Marks per questions/areas of the CRT cases (Full question wording included in Appendix 3)

Categories	Marks case 1	% Total case 1	Marks case 2	% Total case 2
Selection of information	27	37%	18	29 %
Hypothesis generation (diagnosis)	30	40%	30	48 %
Hypothesis testing I (Data interpretation)	17	23%	14	23%
Subcategories (by order in questions)				
Selection of information I (Q5) - Processing information of patient presentation	10	13%	10	16%
Initial information processing (Q7)	3	4%	3	5 %
Early decision making (Q8 + Q9)	10	13%	10	16%
Selection of information II (Q10** + Q11) - Selection of discriminatory questions	14	18%	5	8%
Diagnostic hypothesis I (Q13) - early diagnostic hypothesis base don limited information about the patient	8	11%	8	13%
Prediction (based on previous hypothesis) PE I (Q15)	3	4%	3	5 %
Diagnostic hypothesis II (Q17) - reformulation due to new data	8	11%	8	13%
Decision making (Choose Q18) II - Choose CI accordingly to the DDx (Q13 and Q 17)	5	7%	5	8%
Prediction (based on previous hypothesis) CI II (Q19)	2	3%	2	3%
Hypothesis evaluation/reformulation (due to new facts) I (Q20 to Q25)	6	8%	4	6%
Diagnostic hypothesis III (Q26) - Diagnosis definition and data interpretation	2	3%	2	3%
Diagnostic hypothesis IV (Q27) - justification of diagnosis	2	3%	2	3%
Hypothesis evaluation/reformulation (due to new facts) II (Q32)	1	4%**	-	-
Total possible marks	74		62	

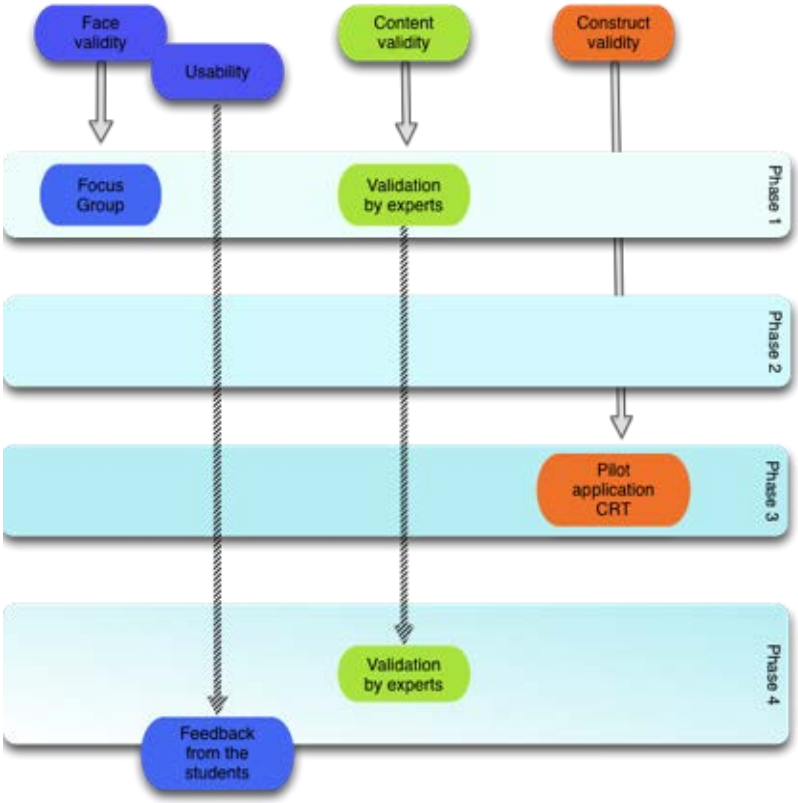
**Only considered in case 1: MI, as this question was a not included in case 2

B.2.4. Validation of the Clinical Reasoning Test (CRT)

Several steps were made to test the CRT's validity. A very small focus group was conducted initially to access students' perceptions of the CRT, its perceived difficulty, its face validity and usability aspects. After that a review by an extended panel of experts was conducted in order to assess the content validity of the instrument. A pilot study was then carried out to assess its reliability. Based on these results some alterations were made to the CRT, with two cases removed, followed by a new review by experts.

Finally during all these processes feedback was constantly asked from the students, although most was given informally in the pilot study sessions. This was crucial and ensured that small problems with answering formats, data entry and access to the test, that occasionally occurred during pilots, could be dealt with appropriately and did not have an impact on the final version.

Figure 6. 3: Validation process of the CRT



Focus Group

Although an email had been sent out to all BMedSci and PBL course students who had just started their second clinical placement (CP2) asking for focus group volunteers, recruitment prove to be extremely difficult.

A small focus group took place on the 3rd of September 2010 with three medical students who had completed one semester of clinical practice, from the University of Nottingham Medical School.

Three female BMedSci students volunteered to take part. The students were asked to answer four CRT cases: a case of myocardial infarction, a case involving diabetes complications (Diabetic ketoacidosis), a chronic respiratory disease case (Chronic obstructive pulmonary disease), and a gastroenteritis by *campylobacter* case. They were instructed to focus on assessing the instrument rather than providing full answers to the questions. Aspects of concern were the quality of the questions (content and format); the value of the instrument (face validity) and the difficulty level. Following a semi-structured focus group script, a discussion was facilitated by the researcher in order to collect their opinions on the aspects mentioned, and whether they had felt the instrument was assessing their clinical reasoning skills. The focus group lasted a total of approximately 45 minutes after the students had answered the CRT cases. The session was not recorded, but the researcher was given permission to take notes of the discussion. The focus group complied with the ethical principles of the research: data collected was anonymous, participants were informed and consented to their participation in this study. A summary of the aspects identified by the participants is provided in Appendix 4.

Based on the feedback from the focus group some changes were made to the CRT cases:

Initial information language was simplified and bullet points added for clarity and bigger fonts were used.

Technical issues were dealt with, and audio recordings were replaced with better quality ones.

A short-answer question was added asking the students if they had any more questions for the patient.

Normal values were added.

Time needed for taking the tests was considered problematic therefore a plan was devised for arranging times that could suit different students, as

well as providing additional rewards for participation, such as a small £10 prize for best results and giving feedback from experts regarding the cases.

Pilot study

Pilot Study: Data collection procedure

After the focus group and after careful reassessment of the CRT cases by the researcher and the experts and a new version of the CRT with four CRT cases was used to conduct a pilot study.

For the pilot data collection a few "Clinical Reasoning Sessions" were arranged. These sessions had a first part where the students answered the CRT Pilot followed by "clinical reasoning tips" given by an expert. The sessions were advertised using both the Network Learning Environment of the University of Nottingham and students' emails. Additionally the researcher made announcements in some lectures and posters were placed in visible places around the Medical School. It was made clear to students that if they could not attend the sessions but wished to take part, then other more suitable times could be arranged for them, although without the expert feedback. Three sessions took place, with a total of 32 students from the 3rd year (year before entering clinical rotations), CP1, CP2 and CP3.

Sessions took place on different days and times in order to try to maximise the availability of students: 2nd and 3rd of September (3 students); 6th (1 student) and 15th of October (16 students), 3rd (9 students) and 4th of November 2010 (3 student) and a final session 22nd of February 2011 (2 students).

Pilot Study: Sample

A total of 32 students took part in the CRT pilot phase, each student was asked to answer the four clinical cases with a total of 116 questions (case questions plus research questions) and three distracter tasks (sudoku puzzles) placed between cases. A wide range of students participated in this pilot study, from 3rd year students to final-years, with the majority of students being either in the first semester of clinical practice (CP1) or in their third (CP2). Our sample was mainly female (71.9%) and mostly were on the BMedSci course (A100 or A300) (16 plus 5, total 21) as opposed to the PBL course. Portuguese students were not involved at this stage. Recruitment of Portuguese medical students for the pilot study proved to be difficult.

Therefore to avoid delays in the pilot study only a UK sample was used. We tried to minimise the impact of this limitation by seeking the active participation of teachers from the Portuguese medical school in the validation and development of the CRT cases and by ensuring the careful translation procedures.

Table 6. 5: Gender distribution on pilot sample

	N	Percentage
Female	23	71.9
Male	9	28.1
Total	32	100.0

Table 6. 6: Year distribution on pilot sample

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3rd year	4	12.5	12.5	12.5
	Cp1	14	43.8	43.8	56.3
	Cp2	11	34.4	34.4	90.6
	Cp3	3	9.4	9.4	100.0
	Total	32	100.0	100.0	

Pilot Study: Results

One of the first observations made, when collecting the pilot data, was that students were not answering all the questions, with an increased number of missing values registered in cases 3 and case 4. For the last two cases the number of completed answers was in fact too low to allow statistical analysis to be performed. From these results and the feedback collected (informally) during the sessions, it was clear that the test length of four cases was too long and that students would get increasingly tiered and disinterested after 45 minutes to 1 hour. Due to the number of missing values in cases 3 and 4, performing a reliability analysis on all cases combined was not possible or recommended (Toutenburg 1998). For this reason, the reliability index was calculated, using Cronbach's Alpha for case 1 and case 2 separately and combined. For this analysis only the case related item were used, research questions and feedback questions were not included in the analysis.

Reliability testing gave a case 1 alpha value of 0.836 and case 2 value of 0.843, which are considered good/very good reliability values. Reliability for case 1 and 2 combined was 0.948 on standardised items (Table 6. 7 to Table 6. 9below) which can

be considered a very good level of reliability (Gliem & Gliem 2003). However it has to be taken into consideration that the large number of items (77) might have an impact on the increase of this coefficient (Tavakol & Dennick, 2011).

Table 6. 7: Reliability case 1

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.836	.935	37

Table 6. 8: Reliability case 2

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.843	.930	33

Table 6. 9: Joint reliability case 1 and 2

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.948	.981	77

Comparison between years

Comparison between means was used to determine the discriminatory sensitivity of the CRT cases (non parametric tests). It is possible to observe that the CP3 students performed better in case 2, closely followed by the CP1s. In case 1 the CP1 students performed slightly better, than the CP3 students. Third years students had the lower mean rank for case 1, while on case 2 it was the CP1 group that had the worst performance. However, no significant differences were found for both case 1 and 2

between the means of students in different years. This might be due to the size of the sample (e.g. only three CP1 and four 3rd year students answered the cases).

Table 6. 10: Comparison between years overall results case 1 and case 2

	Year	N	Mean Rank
Case1 total	3rd year	4	15.50
	Cp1	14	16.82
	Cp2	3	15.50
	Cp3	11	16.73
	Total	32	
Case 2 total	3rd year	4	12.88
	Cp1	14	17.71
	Cp2	3	11.00
	Cp3	11	17.77
	Total	32	

Table 6. 11: Difference between year overall results case 1 and case 2 (Kruskal Wallis Test)

	SMEAN (case1total)	SMEAN (Case2total)
Chi-Square	.221	2.181
df	3	3
Asymp. Sig.	.974	.536

a. Kruskal Wallis Test
b. Grouping Variable: Year

Pilot Study: Implication for development

As said before, a large number of missing values was observed in cases 3 and 4. The high number of missing values found in cases 3 and 4 precluded the performance of statistical analyses concerning these cases. This was a product of the time taken by students to answer the four CRT cases combined, more than 1 hour, leading to a lack of attention and interest in the last cases. This is an important finding concerning the feasibility and size of the CRT. Based on the above results it was decided to use only case 1 and 2, that is the myocardial infarction case, and the diabetes complications (Diabetic ketoacidosis) case, as these two cases did present good to very good reliability indices individually and combined.

The CRT structure was also reviewed in order to identify questions that could be removed in order to minimise the answer time. In order to reduce the size, two criteria were used: a) value of the questions for the aims of the study b) the statistical impact on reliability of the cases. A reduction in the number of questions is expected to lead to a decrease in Cronbach's alpha values (Tavakol & Dennick, 2011); therefore, it was fundamental to ensure that the good reliability of the CRT cases was not significantly affected by the removal of questions.

Treatment questions were identified as questions to be removed. First, because our main focus was to study the clinical reasoning process, focusing mainly on the strategies used to achieve a diagnosis. Decisions regarding treatment and management of patients are not often included in diagnostic/clinical reasoning research, maybe because of their complexity and researchers' difficulty in finding appropriate measures. Second, after analysing the reliability indices of the cases without these questions (tables below) it was observed that the CRT cases would still hold a good reliability index (Reliability Case 1 and Case 2).

Table 6. 12: Reliability case 1 (without treatment questions considered)

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.816	.935	27

Table 6. 13: Reliability case 2 (without treatment questions considered)

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.826	.925	27

To reduce the number of questions of the CRT cases, using the same selection process one other question was identified and removed from case 2 (but not from case 1): "Selection of questions to ask from a list". This question was considered important for our study, as it would reveal the students' ability to choose questions

according to the discriminative value, however a question reliability analysis showed that removing it for case 2 only would not compromise the assessment power of the case.

Table 6. 14: Reliability case 2 (without treatment questions and the selection of questions)

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.796	.918	24

Finally, during this review of the CRT cases, the three questions asking the students about their *"main concerns"* at each moment of the case (after initial presentation, after history questions and after physical examination results), were identified as not sufficiently clear. By inspection of the students' answers to these questions, it was possible to observe that some students would provide differential diagnoses while other would just identify problems. Therefore it was decided to re-write this question into two different types of question. One, placed just after initial patient presentation, asked the students to identify the "main problems". A second was a differential diagnosis matrix, being placed after history taking and after the results of physical examination were given.

Review by experts

As explained before in section 0 of the present chapter, the development of the CRT cases was an iterative process where experts had a frequent and fundamental input into the accuracy of the clinical content in order to guarantee the content validity of this instrument. For feasibility reasons, due to the distance, the researcher meet individually with all the experts in different occasions to discuss their comments and presented others comments and discussed differences when those arisen. This represents a limitation of the current review, as ideally the experts would be able to meet and discuss on several occasions the different versions of the CRT. Although, that not being possible technology allowed for interaction and discussion to take place between experts, apart from the individual meeting with the researcher.

Communications were facilitated by the use of technology, mainly through conversation via email, Skype and the features provided by the Rogo for shared comments on assessments.

Selection of Experts

Selection of experts for this group was made on the basis of the medical schools they were from, ease of access, interest in the project and availability to take part. One key aspect was ensuring that people in this group had a multiple backgrounds, so that the cases would not be biased towards a specific specialty.

The selection of experts was based on the medical schools these experts were from in order to ensure that all medical schools in the study were represented. A second criterion was the level of understanding and knowledge about both the intended outcomes of the school and its curriculum. That is, the experts identified were people that, by their responsibilities in their schools, had enough knowledge about the schools' practices and about what would be expected of their students.

The experts that reviewed the CRT can be grouped into two categories: a 'working group' and a 'additional reviewers'. The first group included two clinical research fellows from The University of Nottingham, a surgeon and a general practitioner, a Professor of Rheumatology also mentor and adviser in this project and a consultant clinical tutor from the University of Coimbra and a medical education expert from the University of Nottingham. These experts were involved in the design of the CRT advising on its content and questions. Additional reviewers included the programme director of the Graduate Entrance Medical school (in Derby), the Director of the Medical Education Unit of the University of Nottingham, and Professor of General Medicine and a Professor of Surgery both in the University of Coimbra. This group was asked to advise on particular crucial moments of the CRT development.

Experts comments: the process

Touchstone (Rogo) was a system that allowed internal reviewers to see, approve and make comments on exam papers. This system was used in order to collect the extended group of experts' opinions on the CRT questions. All the above mentioned people were set up as internal reviewers of the paper and were asked to, during a period of two weeks, to log in to the system and register their comments and suggestions.

Experts comments

The experts approved the large majority of the question on the CRT cases (50 questions, 24 case 1 and 26 case 2), no question was marked as non-approved, and only a few questions (4 from a total of 54) received comments by experts. Questions from the last group[45] will be presented below along with the actions made based on their comments

Question: Can you list the main clinical problems of this patient

One expert commented on the similarity between this question and the previous question, where students were asked to summarise the initial patient information. Based on the pilot study results on this question, and on the fact that this was only considered potentially problematic by one expert, this question remained as part of the CRT cases.

Essential questions, important and not relevant questions

Students were asked to select from a list of possible patient questions, which questions were essential, important, but not essential or not very relevant questions.

The experts pointed out that more opportunities to ask information about the patient history might be needed, it was decided to include a short-answer question asking the students if they had any more questions for the patient and this was included in the final version of the CRT cases.

Regarding the main stem of the question experts identified the need for adding extra information (see below) highlighting that the questions should be chosen by their discriminatory value for the case, the following information was added to question:

"The selection of questions should be done based on their discriminatory value, that being how would they help you in searching for the most likely diagnosis."

Clinical reasoning problem type of question:

In this question students were asked to indicate clinical investigation results that would either support or refute their main diagnostic hypothesis (see below). One of the experts singled out that in fact students could simply suggest one clinical investigation, using it as a way of confirming or refuting their hypothesis (Troponin T Normal: opposes; Troponin T raised: supports). This could also cause confusion in the students and lead to some error in the data.

---

45. One of the comments was related with phrase spelling therefore it was not included here, changes were made accordingly.

For these reasons this question was simplified into a simple short-answer format asking the students to predict two results in the clinical investigations they had previously selected as important. This solution would simplify answers and still be able to access students' ability to predict clinical investigations results.

Figure 6. 4: Clinical reasoning problem type of question.

Note: You can either use the letter above to complete this section as in example 1, or write the results you think will refute or support your hypothesis as example 2 shows. Example 1: I) shows single lesion (pulmonary nodule) in the upper right lung.> 1: refute my hypothesis Example 2: Low potassium >2: support my hypothesis

What results would support and what result would refute your 1st (main) diagnosis hypothesis? (if possible try not to use the same test more then once)

Write here your 1st (main) diagnostic hypothesis at this moment:	
Results (write here the results)	Support or refutes
1	1. This result refutes your main diagnosis hypothesis
2	2. This results supports your main diagnosis hypothesis
3	3. This result supports my main diagnostic hypothesis

B.2.5. Clinical Reasoning Test: Limitations & Conclusions

The results of all the validation tests show that the CRT cases are reliable and gather enough validity to assess clinical reasoning for the purposes of this research.

The CRT cases validation process made use of a multiplicity of methods in order to test both validity and reliability of the cases developed; strategies included were a focus group, a review by of experts and a pilot study (previously presented). However, there are still a limitations that can be identified in this process, and certainly would gain from further research in order t fully validate the CRT beyond the remit of the current research.

One of those limitations relates with the modest size sample of both the focus group (3) and the pilot study (32) meaning that data has to be interpreted very carefully. The focus group had only three elements, although these proved to be helpful and their contribution very meaningful, such a small number clearly constrains the possible interpretations made regarding the face validity of the CRT at this stage. To mitigate its impact on the final results, two feedback questions were included in the final CRT allowing assessment of gather further date on the face validity of the instrument[46].

[46] Question 1: On Scale of 1 (unuseful) to 5 (extremely useful) please rate how useful do you think the CRT type of tests could be for preparing you to clinical practice. Question 2: Would you think online cases, such as the ones you answered and feedback, would help develop your clinical/diagnostic reasoning skills?

This fact demonstrates how difficult it is to recruit volunteers to take part in such studies. According to the participants in the focus group, a possible reason is the students' busy schedules and their geographic dispersion across different cities and hospitals, making it difficult for them to join a session that takes place in the Medical School. In addition the length of the CRT may also have contributed towards this recruitment difficulty.

Another clear limitation is the lack of Portuguese students in the validation sample. Although the careful translation procedures were in place, and teachers from that medical school were often consulted, this is still a limitation of this study. Also although the communications and reviews done by these experts was ensure by making use of technology, it would have been better to be able to organise revision groups. To allow more deep discussions.

The length of a test increases its reliability index, therefore a longer test is more likely to have a better reliability then a test with fewer questions (Tavakol and Dennick, 2011). However the CRT contains multiple types of questions which are usually associated with a decrease in the internal constancy and reliability index (Gliem and Gliem, 2003). An effect of the number of items was observed, but the use of multiple questions did not have a negative impact on the CRT cases reliability. The initial CRT cases were long (37 items case 1 and 32 in case 2), and when combined (a total of 77 items) in fact the reliability increased from good levels (0.836 case 1 and 0.843 case 2) to a level of 0.948. The opposite was also observed, when some of the questions were removed to reduce the time of the CRT cases, however the reliability levels were still good 0.816 for case 1 and 0.796 for case 2.

A finally an important limitation of the reliability testing is the fact that sample collected was to small to allowed for a factor analysis to be done. This additional step requires a sample of approximately ten-times the number of questions in the test, and would allow confirm the internal consistency of the instrument, as well as allow to identify of grouping categories (factors) within that test (Tavakol & Dennick, 2011)

Chapter 7: Comparative Study (Study 2) – Results

"Our ignorance can be divided into problems and mysteries. When we face a problem, we may not know its solution, but we have insight, increasing knowledge, and an inkling of what we are looking for. When we face a mystery, however, we can only stare in wonder and bewilderment" (Noam Chomsky)

Chapter Summary

Past and recent research on different curriculum models seems to indicate that there is a differential effect of PBL on the development of clinical reasoning at an undergraduate level (Chapter 2. Background). Some researchers found this effect to be positive, fostering integration of biomedical knowledge (Hoffman *et al*., 2006; Arts, Gijselaers & Boshuizen, 2000) and even a moderate effect on developing reasoning skills (Koh *et al*., 2008). However, others found this effect to be negative, reporting that students in a traditional curriculum produce better results in terms of the structure and flexibility of their reasoning strategies (Koh *et al*, 2008). Our results add new evidence to this discussion.

The present chapter presents the results of a study (study 2) conducted to investigate the impact of three common types of medical curricula on the development of clinical reasoning in undergraduate medical students. We made use of the methodology and the clinical reasoning test (CRT), described in the previous chapter to assess and compare the outcomes of clinical reasoning. Two cohorts of students in three different medical schools were selected on the basis of their exposure to practice. This allowed us to understand not only the relationship between curriculum types and clinical reasoning developments but also how clinical placements and exposure to real patients contributes towards that relationship.

This chapter will begging by presenting the results of the study, followed by a discussion of the results, and limitations of the study. Finally the main conclusions for the on-going debate on the development of clinical reasoning will be highlighted.

A. Results

A.1 Sample Characterisation:

A total of 281 individual students took part in this study as shown by Table 7. 1 below. Sample used in this study represent 23.7% of the population. A higher percentage of CP1 cohort participated in the study, especially from the BMedSci/BMBS and the PBL medical schools, where almost half of the population took part in the current study. Participation from Portuguese students was lower than for the UK for the CP1 cohort (11.5%) but not for final-year students. Final-year participation was lower than the CP1, as expected due to the particular demands of this year (e.g. final exams, foundation training preparation, clinical rotations).

As explained before to increase participation the individual CRT cases (case 1 and case 2) were made available online as progress tests for the CP1 UK students. Although the majority of students answered both cases (176), some students (80) chose to answer only the first case, whereas others chose only to do the second (23). Table 7. 3 below reports the total number of participants in each cohort, the total number of responses collected by each of the CRT cases and the number of students that only answered CRT case 1 and CRT case 2. Table 7. 2 presents the dates of data collection and the exposure to clinical practice, that is time in months and/or years since start of the clinical phases.

Table 7. 1: Number of students per Medical school and sample

Cohorts	Curricula/Medical School	Total number of students in cohort*	Total number of students that took part in the study**	Percentage of groups in sample
CP1	PBL	91	41 (45%)	15%
	BMedSci/BMBS	249	122 (48%)	43%
	TC	253	29 (11.5%)	10%
Final-year	PBL	90	19(21%)	7%
	BMedSci/BMBS	249	28 (11.2%)	10%
	TC	253	42 (16.6%)	15%
Total		1185	281 (23.7%)	100%

* Approximated values based on the number of admitted students per year based on schools websites
** Total number of students that took part in the study either by answering only one of the CRT cases or both.

As explained before to increase participation the individual CRT cases (case 1 and case 2) were made available online as progress tests for the CP1 UK students. Although the majority of students answered both cases (176), some students (80) chose to answer only the first case, whereas others chose only to do the second (23). Table 7. 3 below reports the total number of participants in each cohort, the total number of responses collected by each of the CRT cases and the number of students that only answered CRT case 1 and CRT case 2.

Table 7. 2: Students clinical experience vs. moments of data collection

	Cohorts	Dates of data collection	Month into Clinical Practice/ time to Graduation
CP1	PBL & BMedSci	04.04.11 to 15.04.11 and 10.05.11 to 24.06.11	Between 2 to 4 month into CP1 training
	TC	31.05.11 to 29.06.11	4 month into clinical practice at year 3 (CP1 level)
Final-year	PBL & BMedSci	01.06.11 to 24.06.11	2.5 years of clinical practice/graduation date July 2011
	TC	31.05.11 to 29.06.11	3 years of clinical practice /graduation date August 2011 (02/08/2011)

Due to this unequal number of students taking each of the cases, and in order to guarantee the quality of the dataset, case 1 and case 2 were analysed separately. Additionally a dataset with the answers only from the participants who answered both cases was built. This dataset was used to test for correlations between the overall results in each of the cases, the case results and other variables in study (collected by research question). This only affected the CP1 cohort in the UK (PBL and BMedSci/BMBS) as they were given the possibility of answering the CRT cases as progress tests.

Table 7. 3: Distribution of sample per CRT cases

Cohort	Curricula/ Medical School	N participants*	N both cases	N Case 1**	N Case 2**	Case 1 only***	Case 2 only***
CP1	PBL	41	18	38	21	20	3
	BMedSci/ BMBS	122	42	102	62	60	20
	TC	29	29	29	29	-	-
Final-year	PBL	19	19	19	19	-	-
	BMedSci/ BMBS	28	28	28	28	-	-
	TC	42	42	42	42	-	-
	Total	281	178	258	201	80	23

* Total number of students that took part in the study either by answering only one of the CRT cases or both. **Total number of answers per case is equals the number of students who answered both cases plus the number of students that answer only one of the individual cases. ***Number of students who answer only to one of the cases.

Gender distribution

The majority of our sample are female students (65.5 %) with male students representing only 34.5%, this is reflected across all the sub-sets (case, case 2 and both cases) of the sample.

Table 7. 4: Gender distribution in the sample per year and medical school

Cohort	Curricula/Medical School	N participants	Gender distribution in sample per cohort	
			Female	Male
CP1	PBL	41	21	20
	BMedSci/BMBS	122	89	33
	TC	29	20	9
Final-year	PBL	19	8	10
	BMedSci/BMBS	28	19	9
	TC	42	27	14
Total		281	184 (65.5%)	97 (34.5%)

Time:

The students took on average 28 minutes to answer one of the CRT cases, and 1 hour and eight minutes (on average) to complete the full two cases.

Table 7. 5: Average times to answer the CRT cases

CRT one case			
Average time	Standard deviation (STD)	Maximum	Minimum
00:27:44	00:12:36	01:08:31	00:11:38
CRT both cases			
Average time	Standard deviation (STD)	Maximum	Minimum
01:08:27	00:34:36	02:35:05	00:13:25

A.2 Descriptive Results

The descriptive results presented below are divided according to the purpose of the questions on the CRT: case questions, research questions and feedback questions.

Case questions

Table below displays the descriptive statistics for the case questions of the CRT cases. It is possible to see by the table above that the mean answer for the final diagnosis is in both cases slightly higher for case 2, with a small standard deviation. The mean of marks per answer also seems to be higher for most of the questions on case 2. An exception is question *3. Course of action to follow after initial patient presentation,* where the mean of case 1 is higher than in case 2. One of the reasons for this

superiority of means in case 2 might be the higher percentage of final-year students answering this case (44%) compared with the ones answering case 1 (34%). Also familiarisation with the test format for the sub-set that answered both cases might be an additional factor explaining better results.

Interestingly it is also possible to observe that only 78 of the 257 (Valid N) students who answered case 1 chose to indicate a second differential diagnosis (after physical examination results) while 256 did it just after history taking. Such a difference was not found in case 2, which might indicate that in case 1 the students were making earlier decisions about the diagnosis, compared with case 2.

Table 7. 6: Means by case and cohorts considered

Year	Course	CP1	CP3
Case 1 (mean)	PBL	44.78	47.80
	BMedSci	40.20	46.26
	TC	42.12	46.82
	Mean-Total	41.55	46.85
Case 2 (mean)	PBL	39.68	43.80
	BMedSci	40.42	44.78
	TC	37.53	42.39
	Mean-Total	39.53	43.46

Table 7. 7: Descriptive statistics for CRT cases questions in case 1 and case 2

Question	Possible marks	Format of answer	Case 1:MI					Case 2:DKA				
			N valid	Min	Max	Mean	Std. Deviation	N valid	Min	Max	Mean	Std. Desv.
1. Summarise the patient information.	10 marks	Short answer	252	0.0	10.0	6.1	3.2	200	2.0	10.0	8.07	1.89
2.List of main clinical problems	3 marks	Short answer	252	0.0	3.0	1.1	1.4	182	1.0	3.0	2.86	0.45
3.Course of action to follow	5 marks	Multiple choice	252	0.0	5.0	3.8	1.9	177	0.0	5.0	3.12	2.28
4.Reason for course of action	5 marks	Short answer	257	0.0	5.0	3.7	1.9	200	3.0	5.0	3.90	0.87
5.What questions would like to ask*	9 marks	Multiple choice	257	0.0	9.0	4.4	2.4	These questions were only included in case 1 to increase the difficulty level.				
6. Additional questions	5 marks	Short answer	249	0.0	5.0	3.8	2.3	199	2.0	5.0	3.95	1.14
7.Differential diagnosis (1)	8 marks	Matrix/short answer	256	0.0	8.0	3.6	3.4	200	1.0	8.0	7.08	1.61
8.3 Results you would expected in the physical examination	3 mark (case 1) 4 marks (case 2)	Multiple choice	252	0.0	3.0	1.3	1.2	195	1.0	4.0	2.62	0.54

9. Differential diagnosis (2)	8 marks	Matrix/short answer	78	0.0	8.0	5.07	2.6	200	2.5	8.0	6.87	1.35
10.5 Clinical investigations	5 marks	Multiple choice	231	0.0	5.0	3.3	1.5	200	0.8	5.0	4.12	0.94
11.2 Results expected in the clinical investigations	2 marks	Short answer	243	0.0	2.0	1.5	1.5	200	0.5	2.0	1.82	0.34
12. SCT (1). How new fact will affect DDx	6 marks	Likert scale	257	0.0	4.0	3.3	2.2	200	0.5	4.0	2.93	0.94
13.Most likely diagnosis	2 marks	Short answer	224	0.0	2.0	1.7	0.7	200	0.4	2.0	1.86	0.40
14.Reasons for diagnosis	2 marks	Short answer	257	0.0	2.0	1.5	0.9	200	0.4	2.0	1.70	0.44
15. SCT (2) how new fact will affect main diagnosis*	4 marks	Likert scale	257	0.0	3.0	0.8	1.1	These questions were only included in case 1 to increase the difficulty level.				

Research Questions

We will now discuss the results from research questions aimed at assessing particular aspects that could help us understand the differences and similarities between the case questions scores. The first three questions aim to identify the moment when students identify a diagnosis or differential diagnosis hypothesis: immediately after patient presentation (listing clinical problem), after history (question 2) or after physical examination results (questions 3). The following question assesses students' confidence in their final diagnosis (4), and the impact of history, physical examination and clinical investigations. Question eight asked the students to briefly describe the pathophysiological mechanisms of the disease they had just seen in the case, aiming to understand the impact of students' knowledge about the disease in their performance. The following two questions aimed at understanding the impact of previous contact with the case with students' performances and the two final questions aimed to understand if the students had or had not used additional resources to answer the cases and how useful those resources were. These last questions were required to control for external influences on their performances as the students answered the CRT in their own time without supervision.

Table 7. 8: Descriptive statistics for research questions

Questions	Research question case 1					Research question case 2				
	N	Min	Max	Mean	STD	N	Min	Max	Mean	STD
1.Can you list the main clinical problems of this patient	242	1	2	1.69	0.67	146	1	2	1.53	0.50
2. What do you think are possible differential diagnosis in this case?	238	1	2	1.21	0.41	137	1	2	1.47	0.50
3.Have these findings (physical examination findings) changed your differential diagnosis?	220	1	2	1.36	0.48	144	1	2	1.58	0.50
4.How confident are you that the diagnosis you have indicated is correct?	196	1	5	4.09	0.76	140	1	5	4.21	0.89
5. How important was the following information in your decision (scale from 1-irrelevant to 5-crucial)										
5.1 History	186	2	5	4.35	0.73	134	2	5	4.24	0.63
5.2 Physical examination	190	2	5	3.62	0.88	135	2	5	3.53	0.83
5.3 Clinical investigations	192	1	5	4.65	0.60	139	3	5	4.78	0.45
6. Knowledge about disease	188	2	4	3.68	0.52	129	2	4	3.76	0.51
7. Have you ever encountered a similar case?	193	1	5	2.34	0.71	139	1	3	2.14	0.82
8.Where you have encountered a similar case	179	1	5	3.14	1.10	89	1	5	3.64	0.92
9. Did you use additional resources of information?	193	1	2	1.08	0.34	135	1	2	1.04	0.19
10. Utility of additional resources of information (1 to 5 scale)	24	1	3	1.17	0.57	4	1	4	1.75	1.50

Feedback questions:

Finally we present the results of feedback questions, which were only answered once by each student, even if the student answered both cases. These questions assess students' perceptions of the CRT cases utility and its contribution for preparation for practice. In summary the students do feel the CRT cases are useful preparation for practice (mean=4.31), that this instrument tests their clinical reasoning skills (mean=1) and all would welcome access to similar instruments in the future (mean =1). These questions were introduced, as due to the small number of participants in the students' focus group, further data was necessary to understand student's perceptions of the cases and to be able to discuss the face validity of this instrument. Results below show a very encouraging picture, that is supported by some emails the students sent to the researcher praising the cases and asking to be included if further cases were developed. However, these were only the students' who volunteered to take part on the study, and this in itself is a source of bias, which will be further discussed in the limitations section of this chapter.

Table 7. 9 : Descriptive statistics for feedback questions

Questions	N	Min	Max	Mean	Std. Deviation
11.Useful for preparing you to clinical practice (1 not useful to 5 extremely useful)	200	1	5	4.31	0.823
12. Would you think cases like this would help develop your clinical/diagnostic reasoning skills? (1 yes; 2 no)	204	1	2	1	0.07
13. Would you like to have access to cases like this in the future? (1, yes; 2, No)	163	1	1	1	0
N	278				

A.3 Results case 1: Myocardial Infarction

Case 1 was a case of a 63 old Polish man with a myocardial infarction. This case had a typical presentation, with back pain inserted as confounding information. The physical examination and clinical investigation results were typical, with an electrocardiogram report describing a clear STEMI elevation. One additional question, a script concordance type (SCT) was added after clinical investigations results to increase the difficulty level of the case (see Appendix 5: CRT case 1 and CRT case 2). The students' results will be analysed below.

Normality testing

A test of normality is required since with small sample sizes the assumption of a normal distribution has to be tested prior to the use of any statistical testing (University of Washington 2010). According to the Shapiro-Wilko test when the sigma value is greater than 0.05 then it can be concluded that the sample is normally distributed. The results below guarantee that the results in case 1 overall are normally distributed.

Table 7. 10: Test of normality for total marks

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
QtotalDDxmean	.044	207	.200*	.995	207	.764

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

In Shapiro-Wilko test any sig > 0.05 indicates a normal distribution.

Table 7. 11: Tests for normality between the two cohorts being studied

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Year	Statistic	df	Sig.	Statistic	df	Sig.
QtotalDDxmean	CP1	.046	139	.200*	.995	139	.929
	CP3	.067	68	.200*	.989	68	.811

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

In Shapiro-Wilko test any sig > 0.05 indicates a normal distribution.

Gender differences:

No significant difference in the students' performances due to gender were found in our sample. Male students, although fewer, seemed to perform slightly better than female students (Table 7. 12 below), however, these differences are not significant (ANOVA table below).

Table 7. 12: Descriptive results CRT case 1 per gender

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Female	164	42.8636	7.09677	.55416	41.7693	43.9579	20.80	62.50
Male	93	44.2613	5.66801	.58775	43.0940	45.4286	28.44	60.90
Total	257	43.3694	6.63822	.41408	42.5540	44.1848	20.80	62.50

Table 7. 13: Analysis of Variance (ANOVA) comparison between genders

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	115.934	1	115.934	2.648	.105
Within Groups	11164.969	255	43.784		
Total	11280.903	256			

Significance level sig<0.05, no significant differences found between groups (female and male).

Years differences:

An individual analysis of variance (univariate ANOVA) was performed to test for difference between the two cohorts being compared. Table 7. 14 and

Table 7. 15 below present the descriptive results and the ANOVA table for this comparison. It is possible to observe significant differences between years for a p of 0.05, with the final-years (CP3) students performing better (mean=46.85) then the students in the CP1 (mean=41.56) cohort. Although these may appear small differences in the means, they represent noticeable and significant differences in the distributions of results (

Figure 7. 1). Final-year students results are, with the exception of a few outliers, concentrated between 40 and 60 points while a higher percentage of CP1 results seem to be concentrated 10 points below between 30 to 50 points.

Table 7. 14: Descriptive results CRT case 1 per year

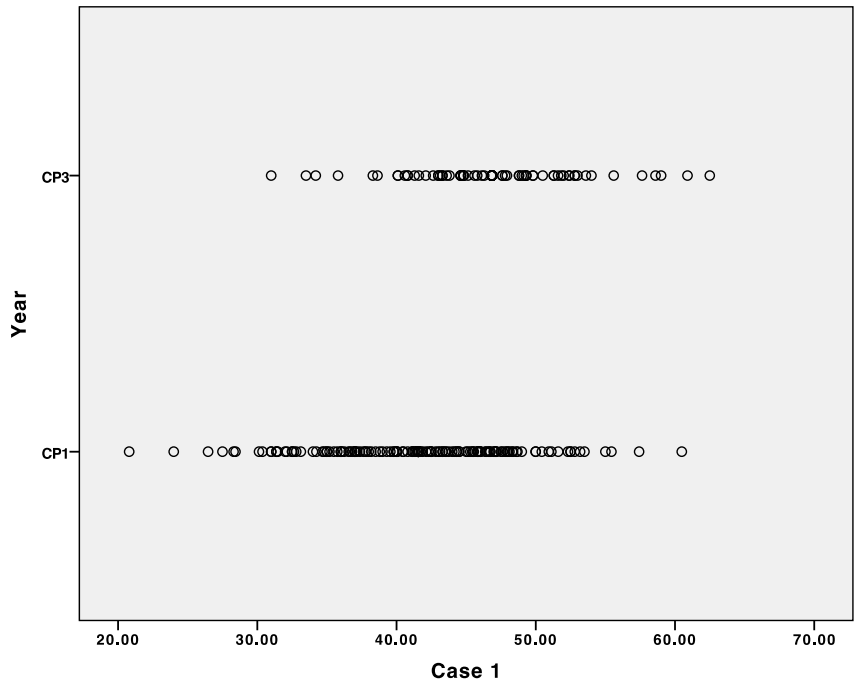
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
CP1	169	41.5552	6.45487	.49653	40.5750	42.5355	20.80	60.49
CP3	88	46.8534	5.52602	.58908	45.6826	48.0243	31.00	62.50
Total	257	43.3694	6.63822	.41408	42.5540	44.1848	20.80	62.50

Table 7. 15: Analysis of Variance (ANOVA) comparison between years

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1624.409	1	1624.409	42.896	.000
Within Groups	9656.494	255	37.869		
Total	11280.903	256			

Sig.<0.05 therefore significant difference between groups variances (CP1 and final-years) are observed.

Figure 7. 1: Scatterplot of CP1 and CP3 results in case 1



Course differences:

The same procedure was repeated for courses as an individual independent variable. Table 7.16 and Table 7. 17 below present the results for the analysis of variance between different medical schools considered in the study. Significant differences between the three medical schools were found, with the students from the PBL course performing better (mean=45.78), followed by the students in the traditional curriculum (mean=44.87) and finally the students from the integrated course (mean=41.50).

Table 7. 16: Descriptive results CRT case 1 per course

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
PBL	57	45.7811	5.96068	.78951	44.1995	47.3627	32.13	60.90
BMedSci	130	41.5034	6.18677	.54262	40.4298	42.5770	24.00	58.60
TC	70	44.8709	7.03162	.84044	43.1943	46.5476	20.80	62.50
Total	257	43.3694	6.63822	.41408	42.5540	44.1848	20.80	62.50

Sig.<0.05 therefore significant difference between courses (PBL, BMedSci and TC) are observed.

Table 7. 17: Analysis of Variance (ANOVA) comparison between courses

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	942.005	2	471.002	11.571	.000
Within Groups	10338.898	254	40.704		
Total	11280.903	256			

Sig.<0.05 therefore significant difference between courses (PBL, BMedSci and TC) are observed.

Based on the above differences it is necessary to understand what the impact of course and year combined was on the individuals' performance. This will then allow us to understand if in fact there are significant differences in clinical reasoning development that can be attributed to the educational environment, that is the curriculum type, and their exposure to clinical practice.

Comparison between Course and Year

Tables below present the results of mean comparison between students from the different curricula. It is possible to observe that PBL students perform better for both cohorts in the analysis although only for the CP1 cohort are the differences between groups significant.

Table 7. 18: Descriptive statistics CRT case 1, by course within years

Year	Course	Mean	N	Std. Deviation
CP1	PBL	44.7725	38	6.41739
	BMedSci	40.1972	102	5.66613
	TC	42.1161	29	7.65977
	Total	41.5552	169	6.45487
CP3	PBL	47.7983	19	4.40914
	BMedSci	46.2619	28	5.73147
	TC	46.8195	41	5.90133
	Total	46.8534	88	5.52602

Table 7. 19: Analysis of variance CRT case 1, by course within years

Year				Sum of Squares	df	Mean Square	F	Sig.
CP1	SMEAN(QtotalDDxmean) * Course	Between Groups	(Combined)	590.587	2	295.294	7.648	.001
		Within Groups		6409.199	166	38.610		
		Total		6999.787	168			
CP3	SMEAN(QtotalDDxmean) * Course	Between Groups	(Combined)	26.805	2	13.403	.433	.650
		Within Groups		2629.902	85	30.940		
		Total		2656.707	87			

sig.<0.05 significant difference between the groups (PBL, BMedSci and TC)

Table 7. 20 below presents the descriptive statistics for the sub-groups considered. It is possible to notice, just by inspection, that for all the medical schools considered the final-year students (CP3) performance is higher. From results in Table 7. 21 it is possible to conclude that there is homogeneity of variance between the students' performance across groups, because the sig value is greater than 0.05.

Table 7. 22 shows that there is a significant difference between students performances based on their year (p=0.016) and course (p<0.005), however, no significant differences between the groups for the interaction between year and

course (p=0.355). An example of a typical interaction between year and course would be TC students' results increase in function of the year, while PBL and BMedSci would not. The lack of a significant effect shows there is no such effect in our data.

These results, along with the previous results shown in

Table 7. 15, indicate that there is strong effect of the year in our data that is not dependent on the course. This is students in the last years perform better than their contra-parts at the beginning of clinical phase for all curricula analysed.

Table 7. 20: Descriptive statistics years within the courses

Dependent Variable:SMEAN(QtotalDDxmean)

Course	Year	Mean	Std. Deviation	N
PBL	CP1	44.7725	6.41739	38
	CP3	47.7983	4.40914	19
	Total	45.7811	5.96068	57
BMedSci	CP1	40.1972	5.66613	102
	CP3	46.2619	5.73147	28
	Total	41.5034	6.18677	130
TC	CP1	42.1161	7.65977	29
	CP3	46.8195	5.90133	41
	Total	44.8709	7.03162	70
Total	CP1	41.5552	6.45487	169
	CP3	46.8534	5.52602	88
	Total	43.3694	6.63822	257

Table 7. 21: Levene's test of equality of error

Dependent Variable:SMEAN (QtotalDDxmean)

F	df1	df2	Sig.
1.821	5	251	.109

Test the null hypothesis that the error variance of the dependent variable is equal across groups a. Design: intercept + course + year + course + year

Table 7. 22: Two-way ANOVA results for interaction between course and year effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2241.802[a]	5	448.360	12.450	.000
Intercept	391655.146	1	391655.146	10875.577	.000
course	300.771	2	150.386	4.176	.016
year	1037.830	1	1037.830	28.819	.000
course * year	74.916	2	37.458	1.040	.355
Error	9039.101	251	36.012		
Total	494673.246	257			
Corrected Total	11280.903	256			

a. R Squared = .199 (Adjusted R Squared = .183)

In summary results from analysis of case 1 show that there are significant differences between the years, with a strong effect from practice in improving the students diagnostic reasoning skills. Also that there are significant differences between courses, if both years considered, with the PBL students performing better then their colleagues in the two other curricula. A two-way analysis of variance results confirmed these differences between courses [F(2, 251)=4.176 p=0.016] and year [F(1, 251)=28.81 p<0.005] but showed there is no significant interactions between the effects of course and year F(2, 251)=1.040 p=0.355.

A.4 Results Case 2: Diabetic Ketoacidosis

CRT case 2 presented a 17-year-old women, with type 1 diabetes with a diabetic ketoacidosis. She had a typical initial presentation: vomiting for two days, diarrhoea, dehydration and generalised weakness. Both the physical examination and clinical investigations showed typical results for the patient.

Normality

As in the previous case a test of normality was required. The Shapiro-Wilko test of normality of the data is presented on table 7.23 and Figure 7.2 below, showing that the data from students' performances on the CRT case 2 is not normally distributed, and that therefore non-parametric tests are required.

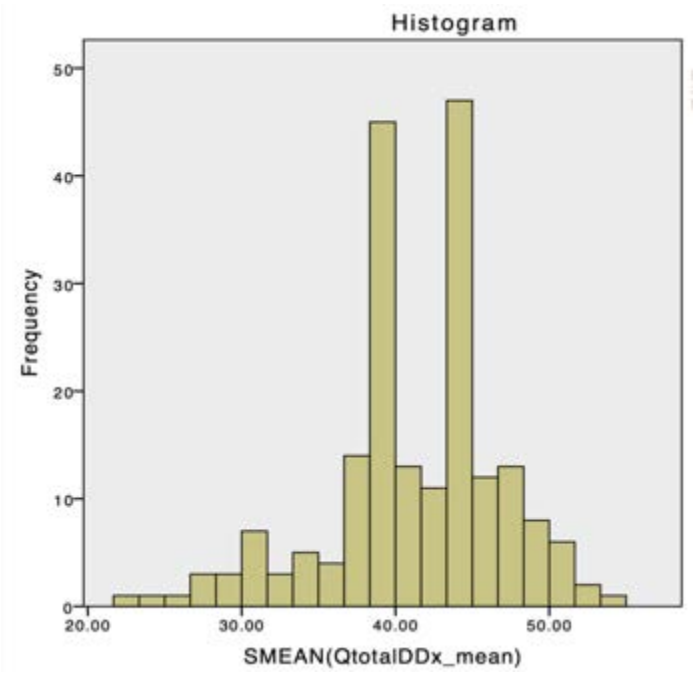
Table 7. 23: Tests of normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SMEAN (QtotalDDx_mean)	.122	200	.000	.960	200	.000

a. Lilliefors Significance Correction

In Shapiro-Wilko test any sig > 0.05 indicates a normal distribution.

Figure 7. 2: Histogram of frequencies students' performance case 2



SMEAN represents the total overall scores of students in the CRT case2

Gender differences:

As in case 1, also in case 2 there are no significant differences (Table 7. 25) between genders were found, although in this case female participants seem to perform slightly better, as shown by the table below.

Table 7. 24: Mean rank distributions between gender

	Gender	N	Mean Rank
SMEAN (QtotalDDx_mean)	Female	133	101.41
	Male	67	98.69
	Total	200	

Table 7. 25: Test statistics for difference between genders

	SMEAN (QtotalDDx_mean)
Chi-Square	.100
df	1
Asymp. Sig.	.752

a. Kruskal Wallis Test
b. Grouping Variable: Gender

In Chi-Square test any sig < 0.05 indicates a significant difference between the groups. No significant difference between genders.

Year differences:

The above procedures were repeated for year, finding significant difference between years (p<0.005), with the final-year performing significantly better then the CP1 cohort (mean 43.46 and 39.53 respectively). This small difference between the averages seems to be due a few outliers in CP3 cohort, that is the few students who scored below 35 points, when that is taken in considerations noticeable differences in the distribution of the results between the two cohorts, CP1 and CP3 can be observed.

Table 7. 26: Kruskal-Wallis Mean ranks per years

	Year	N	Mean Rank
SMEAN (QtotalDDx_mean)	CP1	112	79.06
	CP3	88	127.79
	Total	200	

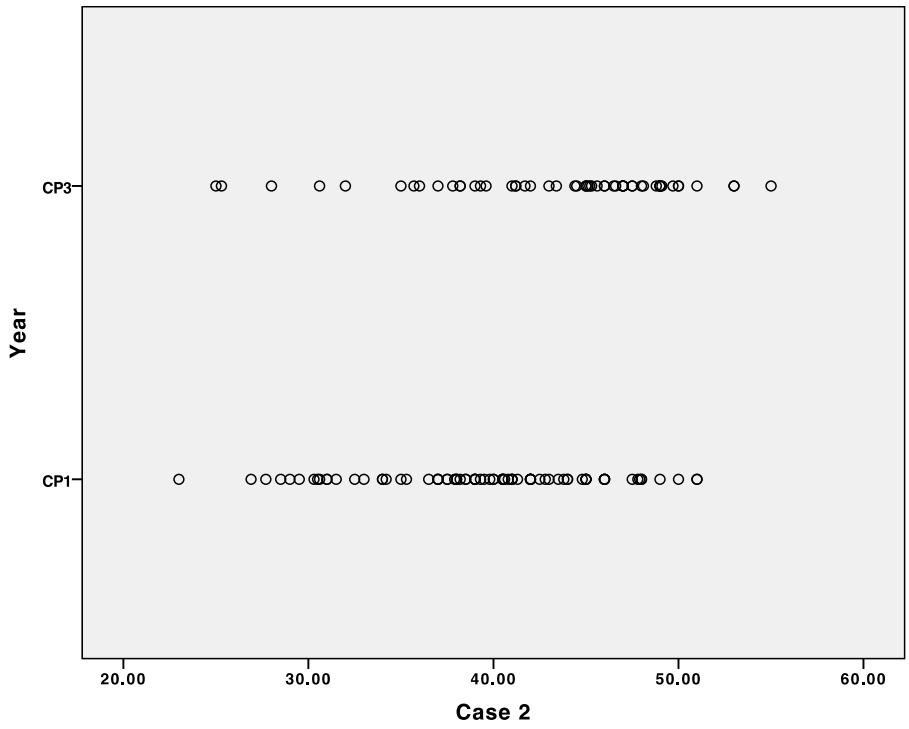
Table 7. 27: Non-parametric test for difference between years (cohorts)

	SMEAN (QtotalDDx_mean)
Chi-Square	35.229
df	1
Asymp. Sig.	.000

a. Kruskal Wallis Test
b. Grouping Variable: Year

In Kruskal-Wallis/Chi-Square test any sig < 0.05 indicates a significant difference between the groups. Significant difference between years.

Figure 7. 3: Scatterplot of CP1 and CP3 results in case 2



Course differences:

In this case the BMedSci/BMBS curriculum group performed slightly better, with the PBL being second, however, no significant differences were found between courses alone (p=0.364).

Table 7. 28: Kruskal-Wallis mean ranks by course (medical curricula)

	Course	N	Mean Rank	Mean
SMEAN(Qtotal)	PBL	39	103.05	41.49
	BMedSci	90	105.48	41.90
	TC	71	92.78	39.72
	Total	200		

Table 7. 29: Non-parametric test for difference between courses (medical curricula)

	SMEAN (QtotalDDx_ mean)
Chi-Square	2.022
df	2
Asymp. Sig.	.364

a. Kruskal Wallis Test
b. Grouping Variable: Course

In Kruskal-Wallis/Chi-Square test any sig < 0.05 indicates a significant difference between the groups. No significant difference between courses

Comparison between Course and Year:

From the between courses study within each of the cohorts it is possible to conclude that there are significant differences in the performance due to the course for the students with little clinical exposure. For this cohort the integrated curriculum students performed significantly better than the others. For the final-year students, although the PBL students perform better, the difference is not significant (Table 7. 31).

Table 7. 30.A Descriptive statistics by course (medical curricula) within year

Year	Course	Mean	N	Std. Deviation
CP1	PBL	39.6758	21	3.65263
	BMedSci	40.4199	62	5.23001
	TC	37.5252	29	5.37100
	Total	39.5309	112	5.11947
CP3	PBL	43.8020	19	7.72182
	BMedSci	44.7818	28	3.67310
	TC	42.3910	41	4.79065
	Total	43.4564	88	5.31054

Table 7. 31: Kruskal-Wallis mean ranks by course (medical curricula) within year

Year		Course	N	Mean Rank
CP1	SMEAN (QtotalDDx_mean)	PBL	21	55.21
		BMedSci	62	62.57
		TC	29	44.45
		Total	112	
CP3	SMEAN (QtotalDDx_mean)	PBL	19	50.71
		BMedSci	28	49.34
		TC	41	38.32
		Total	88	

Table 7. 32: Non-parametric test for difference between courses (medical curricula) within year

Year		SMEAN (QtotalDDx_mean)
CP1	Chi-Square	6.331
	df	2
	Asymp. Sig.	.042
CP3	Chi-Square	4.782
	df	2
	Asymp. Sig.	.092

a. Kruskal Wallis Test
b. Grouping Variable: Course

In Kruskal-Wallis/Chi-Square test any sig < 0.05 indicates a significant difference between the groups. Significant difference between courses in the CP1 cohort, but not in the final-year students (CP3).

A.5 Correlation between cases

Correlation between cases was calculated based only on the results of students that answered both CRT case 1 and CRT case 2 (see section B.1.3 Data collection: Sample from present chapter), number of respondents is shown in Table 7. 33 below.

Table 7. 33: Sub-sample distribution of students who answered both CRT cases (cases 1 + case 2)

Cohort	Curricula/Medical School	Total number of responses to both cases
CP1	PBL	18
	BMedSci/BMBS	42
	TC	29
Final-year	PBL	18
	BMedSci/BMBS	28
	TC	41
Total		176

As it is possible to observe in Table 7. 34 sample sizes are small, therefore it is necessary to test the normality of data distribution before deciding which statistical tests to apply (University of Washington, 2010). This is presented in Table 7. 35 below. It can be observed that Shapiro-Wilko test sigma are all higher then 0.05, therefore it is possible to accept the null hypothesis, that data are normally distributed for both cases and cohorts.

Person's correlation was calculated in order to test for significant correlations between the total mark on CRT case1 and CRT case2. It is possible to see by Table 7. 35 below that there is a significant 2-tailed correlation (r=0.401) for a p=0.01. This significant correlation between the students' performances in both cases supports the idea that the same construct is being assessed. Although these results must be understood with care as this sub-sample has a higher percentage of final-year students than the total of our sample which, as previous results show, have a better performance in the CRT cases. On the other hand, also these students were the ones who chose to take both cases, which can be understood as a sign of interest in the study or the instruments.

Table 7. 34: Test of normality for dataset of students who answered both cases

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Year	Statistic	df	Sig.	Statistic	df	Sig.
case1	CP1	.081	48	.200	.989	48	.931
	Final Year	.107	60	.086	.983	60	.565
case2	CP1	.113	48	.167	.965	48	.156
	Final Year	.077	60	.200*	.981	60	.467

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.
b. Caseanswered = Both

In Shapiro-Wilko test any sig > 0.05 indicates a normal distribution.

.

Table 7. 35: Correlation between students results in both cases for a p<0.01

Caseanswered			case2	case1
Both	case2	Pearson Correlation	1	.401
		Sig. (2-tailed)		.000
		N	109	108
	case1	Pearson Correlation	.401**	1
		Sig. (2-tailed)	.000	
		N	108	139

**. Correlation is significant at the 0.01 level (2-tailed).

A.6 Research questions

A total of 10 research questions were used in each of the CRT cases. The results of their analysis and relationships with the overall performance results will be presented below.

Moments of decision:

This set of questions allows us to understand at what moments in the case the students were feeling confident enough to make key decisions regarding the case. "List main clinical problems" was placed just after the initial presentation of the patient, before the history taking moment. A second set of questions was placed just after history taking and before the physical examination questions and results, and the third questions of this group "have these results changed your diagnosis" were placed after physical examination results were disclosed to the students. No questions were placed after the clinical investigations results were given, as a

question asking the students to indicate a final diagnosis was placed just after the disclosure of these results. Additionally, given the typical results from the clinical investigations, it would not be expected that the students would have doubts about whether or not they would be able to identify a diagnosis.

Case 1:

It is possible to observe from Table 7. 39 that on case 1 question 1, for all the medical schools in this study the majority of students felt they were able to identify the main problems of the patient (159 students from both cohorts), just after the patient presentation, while only 80 students felt they *needed more information* before making the decision. However, in the final-year cohort more PBL and integrated curriculum students chose "need more information". However, when comparing these two groups' (final-year PBL and BMedSci/BMBS) lists of clinical problems (Q2. From case-base questions) with the ones from students in the traditional curriculum, no significant differences were found [$F(2,82,84)=1.261$ $p=0.289$)]. The students from these two groups (final-year PBL and BMedSci/BMBS) who felt they could identify clinical problems did it as well as the traditional curriculum colleagues.

When asked to list the differential diagnoses (question 2), before any results of the physical examination were known, 80% (188/236) of students said they "need more information". Nevertheless, in the final-year this difference is much smaller with only 10 students reporting the need for more information, and within this cohort, the PBL students (8) actually felt they could make a decision at this point compared with (5) that felt they needed more information.

Question three was placed just after the physical examination results were given to the students. This information made 69% of CP1 students' (109 in 158) and 51.7% of final-year students (31 in 60) change their differential diagnosis hypothesis. However, in the final-year cohort only 28% (3/11) of the PBL and 47% (8/17) of the BmedSci/BMBS students changed their differential diagnosis due to physical examination results, while 62.5% (20/32) of the students in the traditional curriculum reported doing so.

Table 7. 36: Frequency distribution per course identification of main problems based on initial patient presentation

Can you list the main clinical problems of this patient * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	Can you list the main clinical problems of this patient	I need more information	10	12	22
		I can make a decision at this point	35	2	37
	Total		45	15	60
BmedSci	Can you list the main clinical problems of this patient	I need more information	21	12	33
		I can make a decision at this point	76	7	83
	Total		97	19	116
TC	Can you list the main clinical problems of this patient	I need more information	13	12	25
		I can make a decision at this point	14	25	39
	Total		27	37	64
Total	Can you list the main clinical problems of this patient	I need more information	44	36	80
		I can make a decision at this point	125	34	159
	Total		169	71	240

Table 7. 37: Frequency distribution per course of ability to identify a DDx after history taking

What do you think are possible differential diagnosis in this case? * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	What do you think are possible differential diagnosis in this case?	Need more information	43	5	48
		I can make a decision at this point	4	8	12
	Total		47	13	60
BmedSci	What do you think are possible differential diagnosis in this case?	Need more information	91	10	101
		I can make a decision at this point	6	8	14
	Total		97	18	115
TC	What do you think are possible differential diagnosis in this case?	Need more information	16	23	39
		I can make a decision at this point	10	12	22
	Total		26	35	61
Total	What do you think are possible differential diagnosis in this case?	Need more information	150	38	188
		I can make a decision at this point	20	28	48
	Total		170	66	236

Table 7. 38: Frequency distribution per course of changes in DDx due to physical examination results

Have these findings changed your differential diagnosis? * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	Have these findings changed your differential diagnosis?	yes	30	3	33
		no	12	8	20
	Total		42	11	53
BmedSci	Have these findings changed your differential diagnosis?	yes	64	8	72
		no	28	9	37
	Total		92	17	109
TC	Have these findings changed your differential diagnosis?	yes	15	20	35
		no	9	12	21
	Total		24	32	56
Total	Have these findings changed your differential diagnosis?	yes	109	31	140
		no	49	29	78
	Total		158	60	218

Case 2:

Question 1 (Table 7. 39) in contrast to the previous case, shows that overall more students felt they could identify the clinical problems. Although more students in the final-years of the BMedSci/BMBS curriculum still felt like they needed more information (11 students compared with 4 that stated they could make a decision).

With questions 2, as shown in Table 7. 40 below, only 47.6% (69/145) of the students said they would need more information before making a differential diagnosis, while 52.4% (76/145) said they would be able to make a decision.

These results show a clear difference between students, in both cohorts, from different curricula, while the PBL (17 in 26) and the BMedSci/BMBS (40 in 60) clearly felt they could decide at this point, a large majority of the students (43 in 50) in both cohorts of the traditional curriculum chose the "I need more information" option.

In question 3, Table 7. 41, overall the physical examination findings in case 2 did not change students' differential diagnosis hypotheses, except for the students with less clinical experience on the traditional curriculum. For 62.5% of students in both cohorts of this school the physical examination results resulted in a change in the differential diagnosis.

Table 7. 39: Frequency distribution per course identification of main problems based on initial patient presentation

Can you list the main clinical problems of this patient(2) * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	Can you list the main clinical problems of this patient(2)	I need more information	9	4	13
		I can make a decision	11	7	18
	Total		20	11	31
BmedSci	Can you list the main clinical problems of this patient(2)	I need more information	21	11	32
		I can make a decision	26	4	30
	Total		47	15	62
TC	Can you list the main clinical problems of this patient(2)	I need more information	9	15	24
		I can make a decision	12	16	28
	Total		21	31	52
Total	Can you list the main clinical problems of this patient(2)	I need more information	39	30	69
		I can make a decision	49	27	76
	Total		88	57	145

Table 7. 40: Frequency distribution per course of ability to identify a ddx after history taking

What do you think are possible differential diagnosis in this case?(2) * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	What do you think are possible differential diagnosis in this case? (2)	Need more information	7	2	9
		I can make a decision at this point	9	8	17
	Total		16	10	26
BmedSci	What do you think are possible differential diagnosis in this case? (2)	Need more information	16	4	20
		I can make a decision at this point	31	9	40
	Total		47	13	60
TC	What do you think are possible differential diagnosis in this case? (2)	Need more information	17	26	43
		I can make a decision at this point	2	5	7
	Total		19	31	50
Total	What do you think are possible differential diagnosis in this case? (2)	Need more information	40	32	72
		I can make a decision at this point	42	22	64
	Total		82	54	136

Table 7. 41: Frequency distribution per course of changes in DDx due to physical examination results

Have these findings changed your differential diagnosis?(2) * Year * Course Crosstabulation

Count

Course			Year		Total
			CP1	Final Year	
PBL	Have these findings changed your differential diagnosis? (2)	yes	7	3	10
		no	9	7	16
	Total		16	10	26
BmedSci	Have these findings changed your differential diagnosis? (2)	yes	12	3	15
		no	35	11	46
	Total		47	14	61
TC	Have these findings changed your differential diagnosis? (2)	yes	15	20	35
		no	9	12	21
	Total		24	32	56
Total	Have these findings changed your differential diagnosis? (2)	yes	34	26	60
		no	53	30	83
	Total		87	56	143

Correlation between cases and knowledge

One of the research questions asked the students to briefly explain the pathophysiology of the disease they had just encountered in the case. These results were used to understand if there was any significant correlation between the students' knowledge about the disease and the case overall as shown in Table 7. 42 and Table 7. 43 below. No significant correlation was found for any of the cases. One possible explanation could be that our knowledge question did not have an adequate level of difficulty or discriminatory value. Those questions reported a mean of 3.64 and 3.75 (from a total of 4 points) with standard deviation of 0.55 and 0.51 for case 1 and case 2 respectively. Other possible explanations can be related to the time of knowledge testing and the position of the question in the CRT structure, and will be further explored in the section dedicated to discussion of results.

Table 7. 42: Correlation between students overall performance on case 1 and knowledge about the disease MI

Caseanswered			case1	Knowledge about disease
Both	case1	Pearson Correlation	1	.177
		Sig. (2-tailed)		.073
		N	139	103
	Knowledge about disease	Pearson Correlation	.177	1
		Sig. (2-tailed)	.073	
		N	103	126

Pearson correlation defined between 0.00 (no correlation) and 1.00 (perfect correlation), values higher than 0.4 considered significant (if sig.<0.05) correlations, values higher than 0.80 considered significant strong correlations.

Table 7. 43: Correlation between students overall performance on case 2 and knowledge about the disease DKA

Caseanswered			case2	Knowledge about disease (2)
Both	case2	Pearson Correlation	1	.066
		Sig. (2-tailed)		.589
		N	109	70
	Knowledge about disease (2)	Pearson Correlation	.066	1
		Sig. (2-tailed)	.589	
		N	70	114

Pearson correlation defined between 0.00 (no correlation) and 1.00 (perfect correlation), values higher than 0.4 considered significant (if sig.<0.05) correlations, values higher than 0.80 considered significant strong correlations.

Correlation with previous experience

One other research question was aimed at testing the impact on students' overall performance of their previous experience with similar cases. Students were asked to state whether the case was entirely new to them, or if they had seen some of its features (but not the full case) before, each option was attributed a value of 1, 2 and 3 respectively.

Case 1 was new to only very few students and between 6.5% to 18.2% said they had never seen the case. Final-year students for all the cohorts had more previous contact with a similar case then the CP 1 cohort (see table 7. 44). For case 2 more students indicated that this case was new to them. For example, in the integrated

curriculum 50% of the CP1 students had not seen the case before, while all of the final-year students on that curriculum had seen that case or some of its features (Table 7. 44).

A significant correlation for a p value of 0.001 was identified for case 2 (DKA) between the overall score of the students and their previous contact with the case (Table 7. 45). However, that was not the case for the first case 1 (Table 7. 46). A possible reason for this might be that both cases were too familiar to all students, small sample sizes and other possible explanations will be presented at the discussion section of this chapter.

Table 7. 44: Frequencies of answers to "Have you ever encountered a similar case?" per year within course

Course		PBL		BMedSci/BMBS		TC	
Year-cohort	Answers	CP1	Final-year	CP1	Final-year	CP1	Final-year
		Column N %	Column N %	Column N %	Column N %	Column N %	Column N %
Case 1	No	12.1%	7.7%	18.2%	6.3%	9.5%	6.5%
	Some features	45.5%	30.8%	42.9%	18.8%	52.4%	51.6%
	Yes	42.4%	61.5%	37.7%	75.0%	38.1%	41.9%
Total		100%	100%	100%	100%	100%	100%
Case 2	No	35.3%	9.1%	50.0%	0.0%	26.3%	6.5%
	Some features	23.5%	36.4%	23.9%	28.6%	47.4%	38.7%
	Yes	41.2%	54.5%	26.1%	71.4%	26.3%	54.8%
Total		100%	100%	100%	100%	100%	100%

Table 7. 45: Correlation between score and previous contact with Case 1.

Caseanswered			case1	Have you ever encountered a similar case?
Both	case1	Pearson Correlation	1	.156
		Sig. (2-tailed)		.109
		N	139	106
	Have you ever encountered a similar case?	Pearson Correlation	.156	1
		Sig. (2-tailed)	.109	
		N	106	129

Pearson correlation defined between 0.00 (no correlation) and 1.00 (perfect correlation), values higher than 0.4 considered significant correlations, values higher than 0.80 considered significant strong correlations.

Table 7. 46: Correlation between score and previous contact with Case 2.

Caseanswered			Have you ever encountered a similar case? (2)	case2
Both	Have you ever encountered a similar case? (2)	Pearson Correlation	1	.418
		Sig. (2-tailed)		.000
		N	122	74
	case2	Pearson Correlation	.418**	1
		Sig. (2-tailed)	.000	
		N	74	109

**. Correlation is significant at the 0.01 level (2-tailed).

Pearson correlation defined between 0.00 (no correlation) and 1.00 (perfect correlation), values higher than 0.4 considered significant correlations, values higher than 0.80 considered significant strong correlations.

Confidence in diagnosis

Finally a question asked students to report their confidence in their diagnosis. Although no significant differences were found on the basis of the curriculum type the students were involved in, there are significant differences between years (cohorts) for both cases (case1, p=0.024 and case 2, p=0.001 see Table 7. 48 below). The final-year students were significantly more confident then the CP1 group, as it is possible to see by the difference in Table 7. 47. Additionally it is possible to observe below that although very few differences are found between courses, on the second case CP1 students from the PBL curriculum are significant more confident (p=0.008) then their peers from other curricula.

Table 7. 47: Absolute frequency of results for years (cohort CP1 and Final-year)

Caseanswered	Year		How confident are you that the diagnosis you have indicated is correct?	How confident are you that the diagnosis you have indicated is correct?(2)
Both	CP1	N	69	66
		Mean	4.0290	4.02
		Std. Deviation	.85700	.953
	Final Year	N	62	57
		Mean	4.3226	4.51
		Std. Deviation	.56610	.685
	Total	N	131	123
		Mean	4.1679	4.24
		Std. Deviation	.74578	.872

Table 7. 48: Analysis of variance between years (cohort CP1 and Final-year)

Caseanswered				F	Sig.
Both	How confident are you that the diagnosis you have indicated is correct? * Year	Between Groups	(Combined)	5.226	.024
		Within Groups			
		Total			
	How confident are you that the diagnosis you have indicated is correct?(2) * Year	Between Groups	(Combined)	10.580	.001
		Within Groups			
		Total			

sig. <0.05 indicate significant differences between years (CP1 and Final-year) considered

Table 7. 49: Analysis of variance in confidence between courses within years

Year				Sum of Squares	df	Mean Square	F	Sig.
CP1	How confident are you that the diagnosis you have indicated is correct? * Course	Between Groups	(Combined)	1.495	2	.747	1.150	.320
		Within Groups		84.475	130	.650		
		Total		85.970	132			
	How confident are you that the diagnosis you have indicated is correct?(2) * Course	Between Groups	(Combined)	8.711	2	4.356	5.178	.008
		Within Groups		67.289	80	.841		
		Total		76.000	82			
Final Year	How confident are you that the diagnosis you have indicated is correct? * Course	Between Groups	(Combined)	.774	2	.387	1.226	.301
		Within Groups		18.308	58	.316		
		Total		19.082	60			
	How confident are you that the diagnosis you have indicated is correct?(2) * Course	Between Groups	(Combined)	.885	2	.443	.934	.399
		Within Groups		25.115	53	.474		
		Total		26.000	55			

sig. <0.05 indicate significant differences between courses (PBL, BMedSci/BMBS and TC) within the two years (cohorts) considered

A.7 Feedback results

In the last part of the CRT cases three feedback questions were included to give the students the opportunity to express how, in their view, this test was addressing clinical reasoning. They were also asked, if they felt this instrument was useful as a tool to prepare them for practice and if they would like to have access to similar cases in the future. The results are presented below. Overall a large majority of the students found the CRT to be a useful tool to develop their clinical reasoning (Table 7. 50 and Figure 7. 4 below), with 46% of the students stating it would be a useful tool for preparation for practice (Table 7. 51 and Figure 7. 5: Distribution of results feedback question . Finally 96.7% of the students would like to have access to cases like those presented in the CRT (

Table 7. 52 and Figure 7. 6). These results are very positive and encouraging however they cannot be interpreted as absolute values without taking into consideration possible selection bias due to volunteer participation and the small sample considered in the study.

Table 7. 50: Distribution of results feedback question 1

Would you think online cases, such as the ones you answered and feedback, would help develop your clinical/diagnostic reasoning skills?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	missing	71	25.3	25.3	25.3
	yes	202	71.9	71.9	97.2
	no	8	2.8	2.8	100.0
	Total	281	100.0	100.0	

Figure 7. 4: Distribution of results feedback question 1

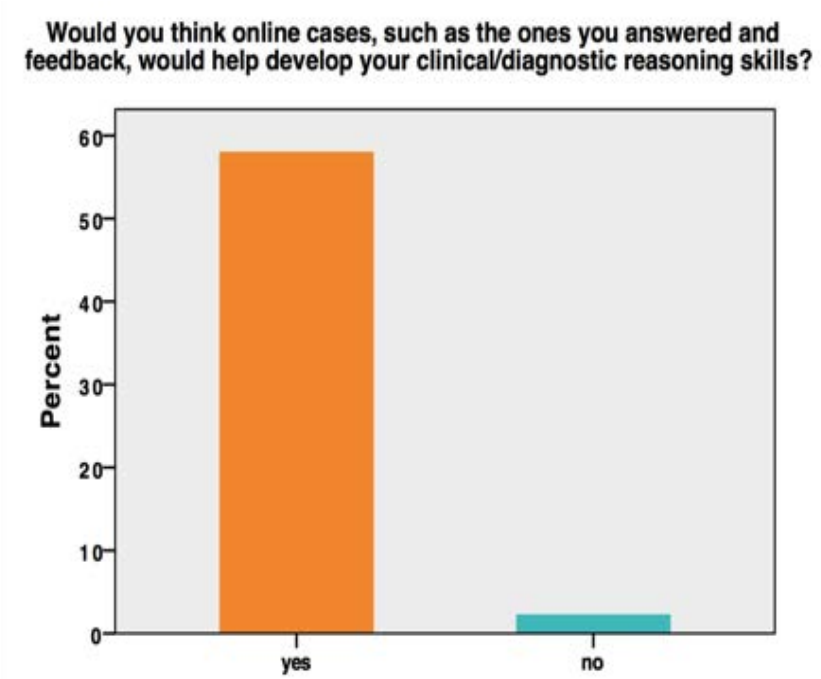


Table 7. 51: Distribution of results feedback question 2

On Scale of 1(not useful) to 5 (extremely useful) please rate how useful do you think the CRT type of tests could be for preparing you to clinical practice

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not useful at all	3	.9	1.2	1.2
	not really useful	8	2.3	3.2	4.4
	not sure about it	32	9.2	12.7	17.1
	useful	93	26.7	36.9	54.0
	extremely useful	116	33.3	46.0	100.0
	Total	252	89.7	100.0	
Missing	Missing	29	10.3		
Total		281	100.0		

Figure 7. 5: Distribution of results feedback question 2

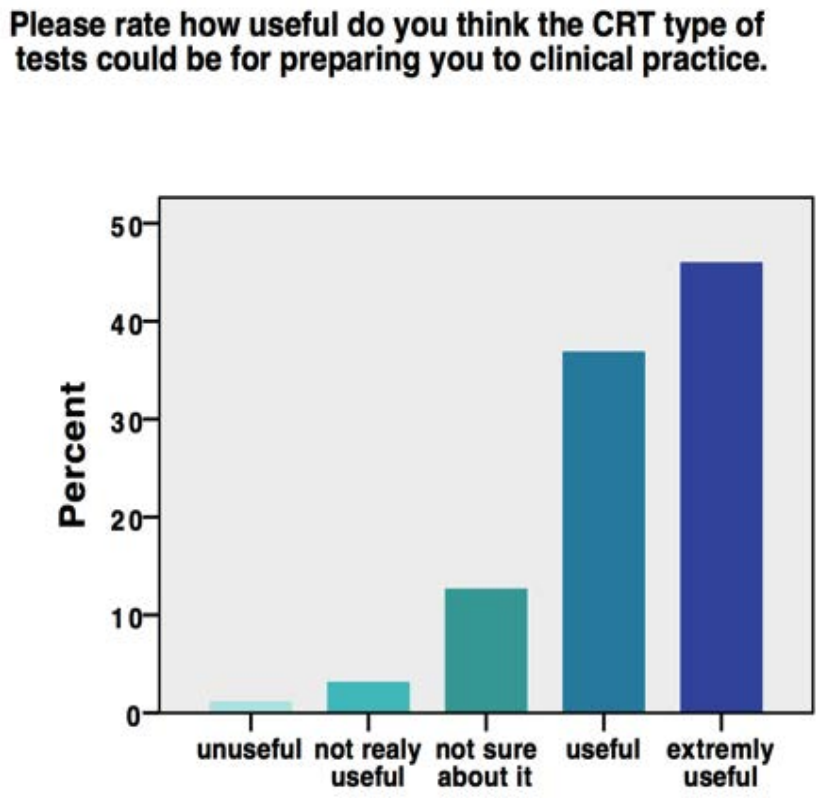
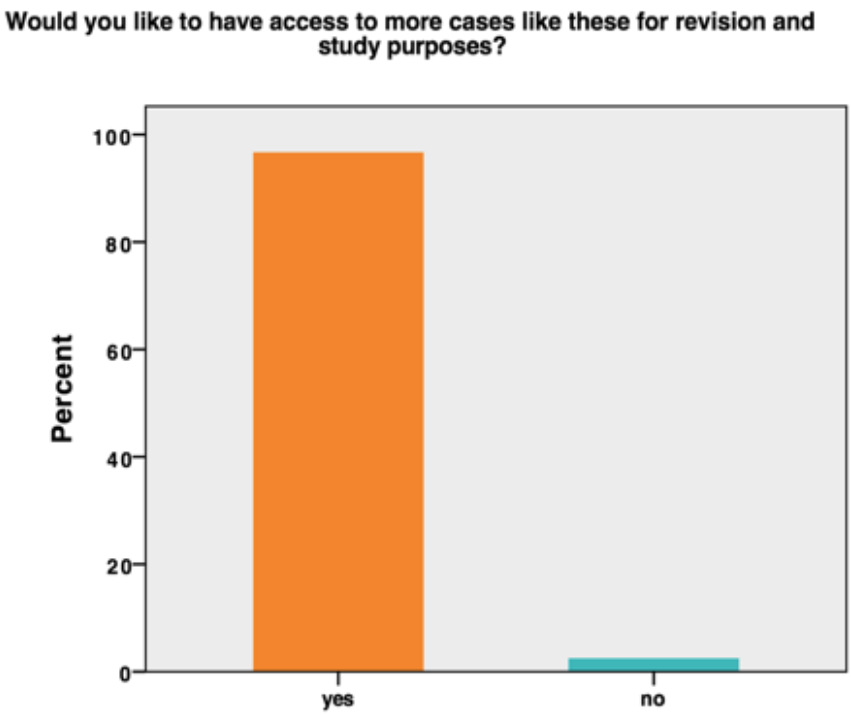


Table 7. 52: Distribution of results feedback question 3

		Frequency	Percent	Valid Percent
Valid	Yes	234	67.2	96.7
	No	6	1.7	2.5
	Total	240	69.5	100.0
Missing	missing	41	14.5	
Total		281	100.0	

Figure 7. 6: Distribution of results feedback question 3



B. Limitations

Identification of the study limitations is a very important part of any research process. It allows for reflection on aspects of the research method and analysis that might have an impact on the acquisition of data and the interpretation of results and therefore needs to be highlighted to enable a meaningful discussion of the results. In the present study limitations can be identified at different levels, mainly for the sampling, data collection and data analysis.

Small Sample and recruitment difficulties

Recruitment of volunteers for this study proved to be difficult. The many demands of the medical degrees, especially bearing in mind both of our cohorts were in clinical practice, spread across different locations and having to dealt with the demands of their clinical placements. Additionally the CRT was designed as an online test, as others have reported in several survey studies (Cook et al., 2000; Vehovar & Manfreda 2001.; Kwak & Radler 2002; Kaplowitz *et al.,* 2004; Truell 2003) this method of data collection is likely to reduce the response rate when compared with paper versions of instruments. Cook *et al.* (2000) reported the mean response rate for 68 surveys was likely to vary between 20% and 39.6%. Kittleson (in Cook *et al.*, 2000) also reports how repeated reminders can, somewhat counter-intuitively, lead to saturation of possible participants and impact negatively in the recruitment. Additionally, the fact that the CRT required on average a half-hour to reply to one of the cases, can also help explain these difficulties.

Our sample size was limited compared with the overall population, representing 21.69%, 16.8% and 14.85% for case 1, case 2 and both cases respectively. Because of the stratification of the sample into two cohorts and three medical schools the subgroups of responses were of a small size. Although testing for normality was performed before using appropriate statistical tests, as is mandatory for small samples, our results still require careful interpretation and no generalisation or determinist statements can be made on those alone. Nevertheless, they provide useful new exploratory information to add to the debate regarding the impact of different curriculum types on the developments of clinical reasoning and highlight some interesting aspects for further research to be carried in this field.

Case specificity: Only two cases

Our instrument was designed, not only to assess students' diagnostic performances but also to provide information about the process, covered mainly by the research questions, focusing on moments of decision, confidence in diagnosis and knowledge about the disease. Additionally the CRT was designed to address all phases of the clinical cycle, from history until diagnosis. Because of these aims it was not possible to include a large number of cases in the CRT, as the testing process would become extremely time consuming. Therefore only two cases were included. According to the literature, clinical reasoning is case, content and context specific (see background chapter) therefore success in one case is not a good predictor of success in other cases, and this should be taken into consideration while discussing the results.

Differential number of answers: Case 1 vs Case 2

A strategy used to improve the number of participants in the study was to upload the CRT cases to UK students as progress tests. To meet the requirements for progress tests used in the schools, CRT cases had to be uploaded as individual cases. Some students chose to answer only one of the two cases. Although the majority of the sample (62.6%) answered both cases, because some students only answered one of the cases answers to CRT case 1 and CRT case 2 had to be analysed individually.

Online without supervision

Students answered the CRT cases online in their own time and without supervision. During the piloting process of the CRT (see previous chapter on development of the CRT) several data collection strategies were tested with the UK students, several different times/days of the week were tested in order to try to find the best moment to schedule the sessions. All these were also presented as options when consulting with the Portuguese students and the study advisers. Neither of the formats proved to be a good enough solution as the number of students coming to these sessions in the UK was always very limited. In Portugal there were several logistical barriers identified to run these sessions with the students, mainly the scarcity of computer rooms available for the researcher to book in the students' locations (clinical placements). Recognising that these students have full schedules and that because they are in their clinical placements and not in the medical school, it was decided to let the students answer the CRT in their own time. To control for additional use of information, two questions were added, one asking the students if they did or did not use information and another asking them to rate the usefulness of information used. Only 5.7% for case 1 and 3.7% for case 2 of the students used additional

resources, no correlation with their performance was found, and all of these students reported the information used was either not useful or that they were not able to access its utility for the case. These results show that the impact of not having researcher supervision of the tests can be considered to have minimum impact on the overall performance of the students.

Time constraints

The CRT cases took on average a half-hour to complete. Not only were there a considerable number of questions, 30 questions for case 1 and 28 for case 2, the students took on average 28 minutes to answer one of the CRT cases, and 1 hour and eight minutes (on average) to complete the full two cases. This added to the recruitment difficulties, although it was observed in the pilot that once the students started to answer the CRT cases they would engage with it. The feedback, from our sample, corroborates this as students rated the CRT as extremely useful and would like to have access to more cases, however, these were responses from those who decided to answer the cases, therefore it cannot be generalised to the all population.

Incomplete responses

Our datasets showed some degree of incomplete responses for participants. Several causes can be identified for this. Firstly, because the link to the online CRT was shared with all the students in the cohorts (1185 students approximately) on several recruitment occasions, it was possible for students to access the CRT and only after trying a few questions deciding to decline participation. However, the system would automatically record this, attributing a participation ID to the student. Secondly, due to the number of question on the CRT it is more likely to have incomplete responses, as this is more frequent in longer tests.

In order to minimise the impact of incomplete responses to the datasets some steps were taken: first data was inspected and entries with only a few answers deleted; secondly the discrete missing values were defined in SPSS so that the software was able identify the cases where very few responses were recorded and treat them as missing values, excluding them from the analysis.

Manual marking and expert model answers

The marking of the short-answer questions was done by the researcher and not automatically by software. As described previously the marking process was designed in order to ensure, as much as possible, the fair consistent marking and feasible marking strategy in face of constraints imposed by its scope, nature of the research

and the time constrains of the experts involved. This however represent a limitation to the study.

Additionally although significant difference between the years based on clinical exposure were found, no student achieved the highest possible score, especially with regards to the script concordance type questions where in a maximum of four marks the maximum achieved was three. This can be a disadvantage associated with using experts' answers as model answers, as students' results would be expectably lower. Also, by adopting this marking strategy there is the danger that some characteristic elements of reasoning at this early stage maybe diluted or missed by the comparison with experts' answers.

Correlation with overall performance in the degrees

As it was not possible to have access to students' full records for reasons that are independent from this research, it is not possible to know if the participants from different groups were equal in terms of their performance in their medical studies, or if, for example, our study was attracting only the best students. Due to the difference between the population numbers and the sample it could be expected that the students taking part in the study are the ones who value clinical reasoning and who are interested in the topic.

In order to try to minimise the lack of information regarding students' profiles a question regarding knowledge about the case was introduced. Students were asked to briefly describe the pathophysiology of the disease they had just seen in the case. This provided a measure of students' knowledge about the disease and it was expected this would have a strong correlation with students' overall results in medical school (Boshuizen, 2003). Therefore, if this study is attracting students with very different overall performances in medical school, significant differences should appear in the knowledge question. Additionally it was not possible to investigate any association between the students' performance and their previous degree, for the Graduate Entry medical school. Therefore extending the results interpretations beyond our sample would be problematic due to lack of representativeness of the sample.

Graduate Entry Students and previous healthcare experience

One of the schools considered in our study is a graduate entry medical school (PBL). Age profiles between Graduate Entry and undergraduate programmes students differ significantly (Garrud, 2011) with the graduate students being older than their peers.

The majority of graduate medical students in the UK would be between 22 and 39 (80.6%) while the majority of undergraduates (UK) would be aged between 17 and 21 (89.3%) (Garrud, 2011). There is no clear indication from the clinical reasoning literature that age alone[47] if not associated with experience in clinical context, would have an impact on clinical reasoning (Kassier, 2009). Though some of these students may have a background or/and work experience in clinical settings that could impact their clinical reasoning ability. However, it must also be noted that also the students in the undergraduate programmes are, nowadays, also expected to have had contact and some experience in the clinical setting. Jiva & Teja (2012) in the Medical Students Application Guide for Undergraduate Applicants state that almost all UK Medical Schools expect students to have healthcare experience, and highlight that as a very important factor in initial assessment of students applications or interview phases. Furthermore, there is some evidence (M Groves et al. 2003) to suggest that although previous degrees may have an influence on students clinical reasoning at an early stage, that is not likely to be the case upon graduation, that is in our CP3 cohort. However the fact that this relationship was not been investigated has to be considered a limitation of the present study. For future studies, an initial question regarding previous healthcare degrees and experience in healthcare settings prior to medical school would allow to better consider these variables impact on clinical reasoning results and avoid this possible source of bias.

Knowledge correlation

As said before this question was introduced on the one hand to minimise the impact of the lack of access to students overall marks in medical school, and to evaluate the relationship between knowledge and the CRT cases results. This question was placed at the end of the case purposefully, as described in the literature (Kassier, 2009). If the CRT was well designed, in order to answer its questions students would necessarily have to mobilise acquired knowledge which would then be available to working memory to answer the question. Our results support this hypothesis, although the level of knowledge demonstrated by the students seems to be superior to that expected, reducing the discriminatory value of this question. Data showed that this question did not have enough discriminatory power, with a large majority of

47. Some studies have looked at age of diagnosticians and accuracy of diagnosis, as a way of assessing expertise. Age is the assessed *per se* but as an indication of higher exposure to clinical practice, engagement in continuo professional development and senior career level. (Eva, 2002b)

students being awarded the maximum marks for the question. Case 1 and case 2 had means of 3.65 and 3.75 (4 maximum points) and only small standard deviation (case 1=0.55 and case 2=0.51).

A few potential reasons can be suggested for this. First the marking system might be too broad, that is a 0 to 4 points for incorrect to correct answer might not be enough to discriminate between highly elaborated answers and simply correct answers. That level of detail was not considered in the marking scheme as our aim was only to understand if the students had satisfactory knowledge of the disease process, and not to classify in detail their levels of knowledge. A second reason is that, as the students answered the CRT in their own time this might have been used as a study tool, along with other study and revising activities. This would naturally lead to an easier retrieval of knowledge from memory (working memory) and therefore better performance on the question.

Differences in samples distributions

Finally differences in sample distributions need to be considered with case 2 answers not being normally distributed this increased the complexity of data analysis. This fact might be explained by the subgroup samples small size along with some incomplete responses. Interestingly when only the answers from participants who had answered both cases were analysed, although the sample size was even smaller, both distributions proved to be normal. One possible explanation is that this shows an effect of familiarisation with the test, therefore students who answered both cases would have been more familiarised with the structure and questions of the CRT. This effect is widely described in the literature (Neuman, 2003; Richmond & Hayne, 1995) with relation to the novelty of tests and tasks. Although the questions types used in the CRT are relatively common within medical education, and instructions for both cases were carefully planned in order to minimise its effects, using multiple screen shots, making instructions clear with examples of questions from the cases. Nevertheless, this is still a possible effect, the inclusion of a test CRT case to allow participants to familiarise themselves with the structure should be considered for future research.

C. Discussion And Conclusions

Clinical reasoning is a complex process and research into the impact of different curricula on the development of this highly important mental process is not consensual (Heemskerk *et al.*, 2008). Although some positive effects of PBL curricula have been identified by some authors (Hmelo-Silver *et al.,* 2008) others have found opposite findings favouring the students in a traditional curriculum (Colliver, 2000). According to learning theories and recent meta-analysis on the effects of the curriculum, it would be expected that the use of active, self-directed, case-base sessions would have a differential positive effect on students clinical reasoning, by comparison with passive knowledge assimilation in traditional curricula (Facione, 2010). It is agreed that experience in clinical settings should foster clinical reasoning if mediated by appropriate planning of outcomes, evaluation, meta-reasoning and reflection (Ericsson *et al.,* 1993). However, how these experiences impact on the relationship between clinical reasoning and the curriculum has not yet been extensively studied.

In order to provide useful information into the above on-going debates, we compared the performance of two cohorts of students, with differential exposure to practice using the CRT in three different medical schools.

As it is possible to see by our comparative analysis of the curriculum, one school's curriculum fits the assumptions of the Flexnerian type of curriculum, where disciplines work in isolation, discipline knowledge is the centre of the organisation structure, and clinical reasoning is only a tacit outcome. In documents from the regulatory body of this school high importance is given to clinical reasoning as a core outcome of undergraduate medical education but few explicitly planned opportunities for its development were found in the school curriculum. Nevertheless, students have the opportunity of experiencing three full years of clinical placements with supervision. Here students are expected to find out relationships between previous knowledge, and practical experience in different specialities.

The second medical school involved in our study fits the description of an integrated curriculum. Here modules created from different basic sciences organised around a theme that is relevant for practice, co-exist with some, less frequent, single theme modules. The student is allegedly at the centre of the curriculum structure, and clinical reasoning is mentioned in some modules, although not directly addressed as

one of the main outcomes of the curriculum. Also here clinical placements are understood as the main learning opportunity for clinical reasoning development, and the students have two and half years of guided clinical exposure.

A third curriculum chosen for the present study is a PBL, trans-disciplinary curriculum. This is an example of a curriculum where knowledge of the different basic sciences disciplines is structured around body systems and presented linked with clinical cases. Throughout the study of these cases the students identify their learning needs, and build their knowledge bases, with the aid of some lectures on relevant themes for the particular cases. Clinical reasoning is an explicit outcome of this curriculum, starting (at the time of this study) to be integrated both formative and summative examinations.

One of the first variables tested for impact on diagnostic reasoning, as assessed by the CRT, was gender. Groves *et al.,* found that in a PBL curriculum the single best predictor of success in clinical reasoning problems was being a female, and this would be a positive predictor of clinical reasoning success. According to these authors the reason for it *"is not readily identifiable."* (Groves *et al.*, 2003, p. 630). Our data do not corroborate that finding. In both our cases or cohorts no significant differences between genders were found. A possible reason for this maybe related with the sample used in the study. Grove's sample was from successive cohorts of one particular graduate entry Medical School in Australia, ours on the other hand was from three Medical Schools and two countries. Any gender bias in one particular schools admission, programme or context would have been diluted in the sample. Another possibility is that these differences are a sub-product of the instruments used.

Additionally, when analysing only the answer from students who answered both CRT cases (1 and 2), a strong correlation between the performance in the cases was found. Nevertheless, more detailed analysis was carried out of each case individually, using all students' answers to each cases (see section: A.1 Sample) in order to test for differences between years (cohorts) and curricula (courses).

For both cases a significant difference was found between years, with the final-year students performing systematically better than the students with less clinical exposure ($p<0.001$). Although the mean scores for case 1 (CP1=41.6 and CP3=46.9) and case 2 (CP1=39.5 and CP3=43.7) have only a small difference, this can be explained by the few identified outliers mainly in the final-years who clearly underperformed compared with their cohort. Additionally, the use of expert's model

answers as criteria for marking the CRT results can also be identified as a factor contributing towards lower results even in the CP3 cohort. Even so, the distribution of the results corroborates the significance of the differences in favour of the CP3 students emerging from the statistical analysis of the variance. This is an expectable result to the fundamental impact of clinical practice in the development of diagnostic and clinical reasoning (Atkinson *et al.,* 2011; Mamede *et al.,* 2008b; Boshuizen *et al.* 1995; Kassirer & Kopelman, 1989; Ajjawi & Higgs, 2008; Norman 2005). It is also a reassuring result, as final-year students graduate to become more autonomous doctors it is uplifting to see from our results that their diagnostic reasoning, as assessed by the CRT, is better than the one of their peers who are still years away from such an important transition.

On case 1 also significant differences ($F(2,254)=11.57$ $p<0.001$) between courses were found, with the PBL students performing better (mean=45.78) than their colleagues from the other two curricula (Integrated mean=41.50, Traditional mean=44.87). As previously the differences in the means score are relatively small but nonetheless significant. A two-way analysis of variance showed that in this case there is no significant interaction between course and year, that is, there is no differential effect of the exposure to clinical practice based on the curricula adopted by the schools. This still needs to be considered carefully as our study is a cross-sectional design with a limited sample. Further research to confirm this effect would require a longitudinal study, following a larger sample of students during at least the 2.5 and 3 years of their clinical placements.

On case 2 results, a significant difference between years was found ($p<0.005$), as it had been in case 1, but no overall significant differences were found between courses. However, when analysing the differences between courses and considering the cohorts students are from, with respect to exposure to practice, it was possible to observe significant difference ($p<0.05$) in the CP1 cohort, with the Integrated curriculum students performing better (mean= 62.57), followed by the PBL students (mean=55.21) and finally the students from the traditional curricula (mean=44.45). In the final-year cohort the PBL students had a better performance (mean= 50.71) compared (mean=49.34) integrated and (mean=38.33) traditional curriculum, but those differences are not significant ($p=0.092$). However, an important consideration must be made here, that is the differences in sample sizes between students from different course was larger than ideally one would like, with the students in the

integrated curriculum representing 55.36% of the sample. This could have had a distorting effect on the results.

Although differences in case results were observed, when analysing only the results from the group of students that answered both cases, a moderated positive correlation (r=0.401) was found between cases (p=0.01).

We did not find, for any of the cases, a correlation between the students' performance in the case and their knowledge of the diseases presented. This seems to contradict previous research when a strong correlation between knowledge and clinical reasoning was found. A possible explanation for our results might be the lack of discriminatory power of our knowledge question as overall performances in this question were too good to be able to discriminate between students who had more or less elaborated knowledge of the diseases. Another possible explanation would also be that although all students had enough knowledge about the diseases (mean=3.64 , 3.75 in 4 possible points STD= 0.55 , 0.5) some of the students had their knowledge organised into elaborated knowledge networks allowing then to perform better in the CRT cases. Additionally, the CRT cases not showing a strong correlation with the knowledge question gives additional support to the fact that what is being assessed here is diagnostic/clinical reasoning and not pure factual knowledge.

The analysis of the moments when the students felt confident to make a decision about the differential diagnosis has shown an effect of clinical exposure and the curricula. Overall, final-year students for the PBL and integrated curricula need less information in order to make their decisions about the case, were less likely to change their differential diagnosis on the basis of new information and were (significantly) more confident in their diagnosis ($p<0.05$) for both cases. In the traditional curriculum group more students, in both years, seem to be require more information to be able to identify a differential diagnosis and are more likely to change it on the basis of new information from the physical examination.

One possible explanation may be that the students in the traditional curriculum are, as suggested by a study from Patel (Patel *et al.,* 1991), more reliant on forward reasoning driven from clinical data interpretations. In this study the authors suggest that students in the PBL curriculum (defined as PBLC in the study) were generating hypothesis earlier and using the clinical investigations data to test and redefine those hypotheses, while the students in a more traditional curriculum were waiting until they had this information (clinical investigations data) to generate their hypothesis by means of pattern recognition (Patel *et al.,* 1991). However, our test was

unidirectional, that is not allowing the students to go back to previous screens and change/review previous answers, therefore it is not possible to confirm if the students in the traditional curriculum, given the opportunity would or not go back and review initial information before making a decision. Testing this possible explanation would be required to allow students to move freely through the test while tracking their movements and choices regarding what information to see and at what stage.

Another possible explanation may lie with the fact that traditional curriculum students, due to the emphases in basic sciences of their curriculum, feel they need clinical investigations data (that is exam results) not only clinical presentation information to make a decision, while others may be more inclined to use the data available form clinical presentations to identify possible diagnostic hypotheses.

While no differences in diagnostic confidence between courses emerge in CP3, for the CP1 group, some mixed results emerged. For case 1 there is no significant difference due to the course but for case 2, PBL students were significantly more confident ($p=0.008$), but not performing better than their peers. For this case the students in the integrated curriculum performed significantly better then PBL or traditional curriculum.

Finally, feedback was extremely positive with a large majority of students considering these cases useful for preparation for practice, expressing the wish to have more examples and considering that these were assessing their clinical reasoning skills. Although recognising the limitations of our study we believe that results here presented highlight interesting aspects of clinical reasoning development and the impact of different curricula.

Overall in our sample there is an effect of clinical exposure leading to a better performance in the CRT, to the use of less information in order to make a diagnosis, a decrease in the likelihood of changing that diagnosis and higher confidence in the decision made. This seems to be the case for students in all the three curricula, with no significant interaction found between the year and the curricula. Although selection bias needs to be taken in consideration, it is possible that only the most interested took part in the study, therefore results could be expectably better than in the overall population, this finding is very positive for all the schools involved.

Attention should be paid to the fact that initial differences existed at the CP1 level, as with in case 1 the PBL students and in case 2 the BMedSci/BMBS students performing

better then their peers, and a question remains to what made those differences disappear? What is happening to make the traditional students 'catch up' with their peers? A response may be time in clinical practice. These students have an extra half-year of clinical exposure when compared with both PBL and BMedSci CP3 cohorts. Would be interesting, however impossible, to see if given that extra time to the PBL and BMedSci undergraduate curricula the differences found in CP1 would persist.

This demonstrates that although with different lengths, PBL 4.5 years, BMedSci/BMBS 5 years and 6 year for the traditional of the programmes, students graduating form these schools do not differ in their diagnostic/clinical reasoning ability. Although clinical placements may be opportunistic and less structured learning environments where clinical/diagnostic reasoning is only a tacit learning outcome (as seen in previous Chapter 3), students are still developing this critical skill. This should be taken as a reassuring finding for the students, the educators and the public. Also, for curriculum developers and stakeholders, a key message might be that pedagogical and curricula choices should be weighted against the available time and resources, and if a more traditional approach to the curriculum is to be adopted then enough exposure to practice should be ensured.

An effect of the case also emerged from our findings, with the results in case 1 and case 2 presenting some differences, mainly with regards to the significance of the differences between courses. This, of course, supports the idea that a performance in one case is not a strong predictor of performance in other different cases, or case specificity.

Finally, we believe that, having a large majority of our sample confirming the face validity of our instruments and considering it a useful way of prepare for practice, along with the results from the validation study developed previously, gives our results an additional credibility (see Chapter 6). Although recognising this is a relatively small sample of students who decided to volunteer for the study, we think that our results do provide interesting and useful information for the debate regarding the impact of medical curriculum design in the development of clinical reasoning which will be explored further in the next chapter.

Chapter 8: Discussion & Conclusions

Science, in the very act of solving problems, creates more of them.

Abraham Flexner

Chapter Summary

The previous chapters presented the results of the two studies conducted in order to understand the impact of different curriculum types on the development of clinical reasoning. In chapter 5 we presented the results of an analysis of PBL sessions, using electronic content analysis and corpus analysis in order to understand how a first-year and second-year group approach and discuss PBL clinical cases. In a second study, in the later chapter 7, we presented the results of a multisite cross-sectional comparative study comparing the diagnostic reasoning of students in the early weeks of their clinical placements and final-year students of three different medical schools,. This study was conducted using a test (clinical Reasoning Test) developed and validated for the purposes of the current research.

In the current chapter, the results from those studies will be discussed in the wider context of the research aims, research questions and existing theories and models of clinical reasoning. Afterwards, conclusions from this research will be outlined and suggestions for curriculum development and further research presented.

A. Summary of findings

In order to be able to discuss the results of the two studies conducted as part of the current PhD research, we provide a summary of the main findings from each of studies (Table 8. 1 and Table 8. 2). These findings must also be considered in light of the philosophical view that underpins our studies: that is pragmatism (Creswell & Clark 2007). The quote from Greene & Caracelli below captures the essence of what was the approach adopted in the current research.

"The importance of context, substantive theory, practical resource constraints and opportunities, and political dimensions of social research as equally important bases for practice decisions. It is time to balance the philosophical, conceptual, practical, and political considerations so relevant to our inquiry" (Greene & Caracelli, 2003 p. 108 in Cameron, 2011, p. 102).

This quote highlights the importance of contextual opportunities in order to be able, by means of scholarship and research, to contribute towards the advancement of what is currently known in a research field. We made use of such opportunities in order to contribute towards increasing the understanding of how clinical reasoning develops in undergraduate medical education and what, as educators and curriculum developers, we can do to foster it. It also highlights that any study is subject to practical constraints, ours is no exception, despite our many efforts, there are necessary limitations to be considered when appreciating our findings. These limitations were presented in the previous chapters and will be highlighted while discussing the results.

Table 8. 1: Analysis of PBL sessions: Content analysis, Chapter 5

Electronic content analysis identified between years: 'Themes' of first-year discussions seem to revolve more around the case information while second-year two discussions seem to be focused more on the basic sciences principles involved in the cases. [Corroborated by the corpus analysis findings (Figure 5.4 and Table 5. 22).
Second-year students use less words in their discussions (Year 1=6115 w/h and year 2=4156 w/h); however, this group uses significantly more technical words but no significant differences were found in the use of common medical words (Figure 5.4).
It is possible to see that there is an increase on the frequency of technical words used by year one students as they progress over time through the different cases. This is confirmed by a significant positive correlation between the clinical case number and technical words shown in Table 5. 11

In the respiratory module (year 1) there is a significant positive correlation showing an increase of the use of common medical words as the students progress through the discussions within this module. Interestingly, in the cardiovascular module are the technical medical words that register such increase over time (Table 5. 14). (Year 1).
For the year 2 group no significant differences can be observed between the cases, or between modules when considering common medical words (Table 5. 18). Although technical words are (statistically) significantly more frequent in the Neurosciences modules with a mean rank of 11 than in the Integrative module with a mean rank of 6 (H(1)=4.5 p=0.034).
We tested for a linear increase in the number of technical and/or common words within each module; however, in the year two group neither show significant correlations (Table 5. 20).
Year 2 use more questions than students in year one (Figure 5. 7, Year 1 students use more explanations and reasoning (combined) than year 2 (Figure 5.8)

Table 8. 2: Comparison of diagnostic ability in the three curricula Chapter 7

The curriculum model had an differential impact on students with few exposure to clinical placements (CP1) diagnostic/clinical reasoning as measured by the CRT
For case 1 PBL students showed significantly better results, while for case 2 the integrated curriculum students performed significantly better.
On the final-year cohorts, no significant differences were found between the students in the three the curricula, as measured by the CRT.
Final-year students performed significantly better than CP1 students.
Overall final-year students from the PBL and integrated curriculum need slightly less information, than the CP1 peers, in order to make their decisions about the case. These students were also less likely to change their differential diagnosis on the basis of new information and were (significantly) more confident in their diagnosis ($p<0.05$) for both cases.
Students from the traditional curriculum, at both levels seem to require more information and change more their differential diagnosis on the basis of clinical investigations information.
No correlation between the case results and the students' pathophysiological knowledge of the disease was found
No correlation between the student confidence and their score in the CRT cases was found

B. Discussion

The research presented here aimed to:

investigate the process of development of clinical reasoning at early stages of the medical undergraduate curriculum; understand the impact of different curriculum types on the development of students' clinical reasoning strategies at the undergraduate level; and understand the impact of clinical experience/exposure on the products of students' clinical reasoning strategies.

In order to achieve these aims, iterative, cyclical approach to research was employed, with a multiplicity of methods. Several instruments were used, and several data collection episodes took place. An analysis of the three curricula informed by the principles of comparative education (Bereday 1964; Bray et al. 2007) was conducted. Two main empirical studies were conducted. One looked at how first and second-year PBL students discuss clinical cases, by applying a methodology adapted from linguistics: corpus analysis. A second study was a cross-sectional study conducted in three medical schools (in two countries), using an instrument developed for the purposes of this research. These studies have been presented previously and discussed separately in their respective chapters.

The results and conclusions of these studies should be understood as complementary in providing answers to the original research questions. Therefore, in the present chapter the discussion of the results of the two empirical studies will be guided by the research questions. It is the combination of perspectives between these studies, looking into the process of clinical reasoning and simultaneously comparing outcomes, that provides unique and distinctive value to the present research, hopefully making a significant contribution to the body of knowledge on clinical reasoning and its development.

According to encapsulation theory, as presented by Boshuizen, students start to build their networks of knowledge by clustering detailed biomedical concepts together, based on their connections under higher level technical concepts (Boshuizen 2003; Rikers *et al.,* 2002; Hobus *et al.* 1987). Boshuizen argued that during the first years of medical school (or biomedical sciences years in traditional curriculum) students should be at a pre-encapsulation stage; that is, they should not yet have enough medical knowledge available for activation and application to the reasoning process.

Therefore, the clinical reasoning process at this stage is often very descriptive, lacking the necessary synthesis and often supported by notes and visual representations of concepts (e.g. drawings) (Boshuizen 2003). According to this model, as students gradually build their knowledge databases, their organisational structures evolve to accommodate the new knowledge. This leads to cluster formation and encapsulation (Boshuizen, 2007). A consequence of knowledge encapsulation would be an increase in the use of concepts that result from joining other concepts together, like "micro-embolism", or "aortic-insufficiency" (Boshuizen, 2003 p. 13). Meanwhile the clinical reasoning process becomes faster, with decreasing cognitive effort, but still following a step-by-step approach between networks of concepts (Rikers *et al.* 2002; Schmidt & Boshuizen 1993).

Our data seems to support this idea. Electronic content analysis showed that second-year students case discussions showed a more complex structure, focused on technical-medical concepts (e.g. pH), than the first-years. This was confirmed by the corpus analysis, with second-year students using significantly more technical than common vocabulary. Also, the number of words per hour of the first-year group discussions was almost much higher than of the second-year group (Year 1=6115 w/h and year 2=4156 w/h) meaning that the students are talking much more, and in a much more scattered way. As Boshuizen said, this is because clinical reasoning at this stage *"lacks the necessary synthesis"* (Boshuizen 2003). This seems to support the idea that second-year students are using more encapsulated knowledge, and are therefore able to use much fewer words and be much more efficient in their approach to clinical cases.

Nevertheless, it is important to note that in Boshuizen's model, encapsulation happens later than the second-year medical school when students have been exposed to substantive time in clinical practice environment (Boshuizen, 2007). A reason maybe that this view is linked to the traditional medical curriculum (Bordage, 1999). It is now consensual that PBL fosters integration of knowledge by exposing students to more clinical cases, which would be expected to foster encapsulation, with several reviews of evidence supporting this claim (Koh *et al.,* 2008; Norman & Schmidt , 2000b, 1992). Furthermore, a study by Patel *et al*., also showed PBL students seem to produced higher quality explanations using biomedical information for clinical problems (Patel *et al.,* 1991), which could be an indication that these students are developing earlier and/or more elaborated knowledge encapsulation.

The above explanations suggest that the differences in language, and in the use of questions and explanations, found between years one and two in our first study, can be the effects of an early encapsulation process being accelerated by the early exposure to clinical cases, and feedback in an educationally planned and purposeful way (deliberated and reflective practice). Moreover, this result would help explain the ones from Hmelo-Silver (1998). This author compared the pathophysiological explanations of first-year PBL and not PBL students, and found PBL students were providing more complex explanations and had a superior ability to transfer reasoning strategies across learning environments (Hmelo-Silver, 1998) which could be a consequence of some degree of early encapsulation of knowledge.

Additionally, an interesting finding was that second- and first-year students differ in the way they approach the cases. Overall, second-year students use more questions while first-years use significantly more explanations/reasoning[48]. These results can be seen as supporting the early encapsulation hypothesis mentioned above, as first-year students would use more descriptive and less connected information, therefore needing to verbalise more their explanations and reasoning in case discussions. However, this alone does not explain the difference in their use of questions. An alternative explanation might be that second-year students start to make more questions because these students are making attempts to discriminate between possible lines of reasoning, more than just explaining what they know about the case. This is a skill often attributed to experts (Papa et al. 1990). However, these students have limited experience, maybe their exposure to clinical cases in PBL discussions (reaching a maximum of 70 cases in the first 2 years of the PBL course) can be fostering the development of these skills.

Another possible explanation might be the lack of sensitivity of our analysis for such a complex process. Another might be that no differences should be expected in the frequencies of questions, but rather in the quality and effectiveness of questions asked.

Nevertheless care should be used to interpret these finding as it is important to consider that differences in the group engagement with the cases and the PBL process, differences in the students backgrounds and individual characteristics, may impact verbalisations and case discussions. Schmidt and colleagues have shown on

[48] We used a combination of reasoning and explanations 'markers' to compare the groups discussions, as shown by our validation study that the later two were very closely related (Chapter 4).

several occasions that many factors impact the group PBL group processes. Commitment to the group learning, attendance to the PBL groups, the quality of the problems, students' previous knowledge and tutors' expertise are some of the suggested (Berkel & Schmidt 2000; Schmidt & Moust 1995; Schmidt *et al.*, 1993). Although both our groups had a clinical and a non-clinical facilitator, and seemed to have a similar student profile, with regard to their GAMSAT scores and to some extent backgrounds, only a larger study with more groups, cases and with a longitudinal design would allow to further confirm if indeed PBL cases are in fact, as our data seem to suggest fostering some knowledge encapsulation.

Based on the aforementioned findings it could be expected that students in the PBL curriculum would perform better than students in a traditional or integrated curriculum in the CRT cases, at least for the cohort without much clinical exposure (CP1). Our results only partially support this hypothesis, showing a positive effect of PBL curriculum on CP1 cohort and case 1 but no clear positive effect of PBL when for both CRT cases considered. These results will be discussed in more detail along with the results of the other similar studies described in literature in the paragraphs below.

Research into clinical reasoning, mainly the work conducted by Groves *et al*., (2002), identified gender as a predictor of positive performance in clinical reasoning, as measured by the clinical reasoning problems. Our results do not support this claim, as no gender differences were found in either CRT cases or cohorts in the study. Nevertheless, our sample is clearly different from the one used by Groves. In their studies only PBL students from one single medical school participated, while ours included students from other curricula and three different medical schools. Also, our instrument, the CRT, was different from the clinical reasoning problems (CRP) used by Groves *et al*., The CRT has different types of questions, as opposed to only one used in the CRP. We also used long cases, not short vignettes as the CRP does.

It is now consensual that knowledge is a fundamental basis to clinical reasoning, (Eva 2005; Norman 2005; Rikers *et al.,* 2002; Boshuizen *et al.,* 1992; Schmidt *et al.,* 1990; Schmidt & Moust, 1995). Our findings show no clear positive correlation between the students' performances and knowledge of the diseases tested in the CRT cases. Our findings suggest all students had a good knowledge of the pathophysiology of the diseases involved in the cases, and even so some were performing better than others in the CRT. This may mean that the CRT cases assess something more than just

factual knowledge. Or it might be an effect of the lack of discriminatory and difficulty value of the marking scheme, awarding too many students too high a score, and therefore making it impossible to see possible differences.

The results of the CRT application show that for the cohort of students with little clinical exposure, CP1, significant differences can be found between the students from different courses ($p<0.005$). PBL students performing overall better than their peers for case 1 (mean =44.78) and students from the integrated curriculum showing the best results for case 2 (mean=40.42 $p=0.042$)[49]. However, both groups (CP1 and final-year) considered a significant positive effect of PBL for case 1 (mean=45.78 $p<0.05$), but no significant differences for case 2 based on the course, although the students from the Integrated curriculum seem to have slightly higher mean scores.

Schmidt *et al.* (1996) also compared the effects of three (traditional, integrated and PBL) medical curricula in the clinical reasoning of undergraduate medical students. These authors used one of the largest samples reported in similar study cases of students from five cohorts of students (from the 2nd until the last (6th) year of medical school) across three different Dutch medical schools, using a larger sample of case (30). These authors found diagnoses from PBL and Integrated curricula students were more accurate than those from the traditional curriculum, with no overall difference between them (integrated vs PBL). Our finding of the CP1 cohorts partially support the ones of this study, but when both CP1 and final-years are considered our results differ from those of Schmidt and colleagues. Some possible explanations for these differences can be identified. Firstly, there are differences in the samples used.

In addition Schmidt *et al.* used five groups in each curricula, from second-year to the sixth year, while in our study we just compared students at the beginning of clinical practice and by the time of graduation. The uneven overall sample distribution in our study, with a high percentage of CP1 students, mainly from the integrated curriculum (43.42%) is problematic, as PBL and the traditional curricula are underrepresented. As the law of large numbers (LLN) implies the likelihood of having cases that do not conform with the norm (e.g. are much better, much worse) is higher in smaller samples (Durrett 2010). Also, it is important to note that different assessment tools were used. The CRT score is not based only on a diagnosis, but on a series of decisions made in various moments of the 'clinical cycle'; the CRT initial information

[49] Due to differences in the sample cases 1 and case 2 were analyses individually.

was presented as an audio recording of a simulated patient, the 30 case vignettes were written summaries of signs and symptoms similar to textbook cases. Still both were case based, both focused on diagnostic reasoning and both limited students' freedom to see only the chosen results in order to avoid consequential error. Also we have used two cases and not 30, which can mean that our results are in fact a product of case-specificity and differences in the content of the cases, more than differences between the curricula. However, both our study and Schmidt's are cross-sectional; less problematic in Schmidt's due the large sample size, but it still means that many confounding variables related with individual and contextual differences can influence the results. Longitudinal studies with repeated measures over a longer period of time would be necessary to confirm these results.

A suggestion presented to explain Schmidt *et al* findings, that would also contribute to explain our findings, is that in fact integrated curricula, as well as PBL foster integration from basic and clinical sciences, and that factor, more the self-directed nature of PBL would be responsible for superior performances in diagnostic tests (Schmidt et al. 1996).

Our results also show a strong positive effect of clinical practice for all involved, with no interaction between the year and course, with final-year students performing significantly better ($p<0.001$) than the CP1 cohort for both cases. This positive effect of clinical exposure has also been found by Schmidt in the aforementioned study (Schmidt et al. 1996) and also by a more recent study by Goss *et al.*, (2011).

However, our results show that, although at the CP1 level significant differences were noted between courses, in the final-year cohort no such differences emerged. It is possible to read from these results that whatever differences exist between groups when students start their clinical placements, by the time these students reach the graduation point no differences can be found between curricula groups. While this is a reassuring finding, a question remains about why these initial differences disappear. These results seem to be consistent with the weak effects of PBL in critical thinking and problem solving skills after graduation reported in the systematic review by Koh *et al.* (2008b). However it is not consistent with the results of Goss *et al.* (2011), also reporting a clear development effect of clinical practice, with no interaction between clinical exposure and the curricula (Traditional vs PBL), but when considering the results of students at the beginning of clinical practice and 'senior' students, a positive effect of traditional curriculum emerged. On the other hand the

overall results from Schmidt *et al.*, (1996), mentioned before, showed significant differences were found for final-years in favour of PBL and Integrated curricula, opposing Goss' results. However, it is important to note that the study of Goss *et al.* was assessing diagnostic reasoning by the diagnostic thinking inventory (DTI) (G Bordage et al. 1990) which is a self-reported measure of self-perception, not diagnostic ability performance, while Schmidt used clinical vignette more similar to the CRT cases.

It is important to note that the curricula compared in our study differed in the amount of time of dedicated to clinical placements. All the curricula used in the Schmidt study lasted for six years. In our study only the traditional curriculum was that long. The PBL lasted less one and a half years (4.5 years total) and the integrated was one year shorter (5 years). This may have an impact on the results as if longer degrees mean more exposure to patients, that is likely to lead to and improvement in the students' clinical reasoning skills (Bowen 2006; Norman & Eva 2005b; Eva 2005b). It would be interesting to know whether, if given the extra time to our PBL and Integrated students, bearing in mind that no two curricula are alike (contextual differences), the final results would be similar to those of Schmidt. This hypothesis is however impossible to test, therefore we should make attempts to explain our results, and those were that clinical practice had a developmental effect on students' diagnostic ability and although differences at an early stages may favour PBL and Integrated curricula at graduation no significant differences between the groups can be found based on the curricula as measured by our two CRT cases.

Some limitations of our study may also account for it. As mentioned before we used only two cases, therefore no generalisations about student performance in order than these two cases can be drawn. Also final-year represented only 31.67% of our sample, with only 19 of those being from PBL, which was clearly underrepresented in our final-year sample. This may explain why although all curricula groups seem to clearly improve with practice in a similar way (interaction absent between year and course), in the final-year group no differences were found. Also, it is possible that the cases difficulty level and the marking scheme based on experts model answers were lacking the ability to discriminate effectively between the reasoning skills of final-year students. Additionally the CRT and Schmidt cases are 'written' assessments, and do not account for complexity of the context, situation, interactional and

metacognitive factors involved in real clinical encounters, therefore their realism is limited, as should be the interpretation of its results.

It is plausible, that CP1 PBL and Integrated curricula students are more familiar with this type of presentation of information. PBL cases and Schmidt's vignettes, and the CRT cases, are to some extent similar. Also, it is possible that these are more commonly used in small group sessions and/or lectures in the integrated curriculum than in a more traditional one, where integration between basic and clinical sciences is less of a priority. Therefore, these students would be less trained to identify relevant clues from these presentations. And with exposure to real patients and the complexity of clinical environments these differences would fade.

Also a possible explanation is that clinical practice has greater impact on students from the traditional curricula. These students are not exposed to clinical cases on a regular basis early in the curriculum; therefore they might feel under greater pressure to respond adequately to the demands of clinical practice. This compensation effect may mean that students make an extra effort to learn from practice and consequently to more meaningful experiences, enhancing their learning at this stage more than the one from integrated and PBL students.

Another possible explanation can be one put forward based on the study of Prince *et al*. (2000). These authors conducted a focus group of PBL medical students. The author argued that PBL students often find practice harder because they need to *think "the-other-way-round"* from what they are used to in PBL cases (Boshuizen 2003; Prince et al. 2000). This means students in PBL cases usually look for explicit single clues (chest pain) that will lead to a diagnosis, and in practice those clues often do not exist. A diagnostic hypothesis is based more on a combination of several different findings than on a single key finding (Boshuizen, 2003). This would imply that someone trained in PBL would find clinical practice more challenging than their colleagues, and have more difficulty integrating learning derived from exposure to practice.

Additionally both PBL and Integrated curricula in our study are organised by blocks/modules related with body systems, while the traditional curriculum is organised by scientific disciplines. The organisation by blocks has been noted as a potential inhibitor of students' ability to explore the problems widely. These students would be organising their knowledge and developing reasoning strategies that would work within a specific block, and those may not be transferable to other systems. Also

the block organisation reduces complexity of problems, by allowing the students to focus on one body system (Sefton et al. 2008). While the traditional discipline-based approach requires the students to understand basic principles and their application to a wider rage of systems and interaction between them. For example students learn their pharmacology principles and are expected to know those principles and to be apply that knowledge to cardiovascular diseases, dermatological problems and psychiatric conditions equally. And very importantly all is assessed together, within the pharmacology discipline. This can explain why in our study for the CP1 students the traditional curriculum students were outperformed, as their basic knowledge is less integrated with clinical aspects and they may lack contact with clinical cases, but after exposure to clinical practice, which addresses this gap, they perform as well as their peers.

There is also the possibility that teaching in the clinical setting might reflect more traditional/science-based models of biomedical knowledge, which favours students in integrated or traditional curricula being more aligned with these curricula than with the PBL. Prince *et al*., conducted focus groups with PBL students during their transition from preclinical to clinical years, and found that students felt very insecure about their knowledge, especially of biomedical basic sciences, including anatomy and pharmacology. They suggested this might be a sign of a lack of integration between these "two worlds". Medical education and curriculum development have been areas of great innovation and change in medical education, while teaching in clinical settings has probably been one of the least developed areas. (Prince *et al*., 2000). The comparison of the three curricula (chapter 3) supports this idea – the large majority of differences between the curricula are focused on how the basic sciences are learned (before clinical practice), but organisation and educational practices in the clinical years are similar.

Finally, there is a possibility that all the above-indicated reasons may be happening simultaneously and that would mean that in fact the absent interaction between year and curricula found in our study, is because students from different curricula are gaining different skills from clinical exposure (compensation effect).

Research into clinical reasoning and the curricula has also looked at possible differences in clinical reasoning strategies due to curriculum types. A study by Patel *et al.* (1991) compared the results of a conventional or traditional medical curriculum with the results of students in a PBL curriculum, and concluded that the PBL

curriculum revealed a more *"backward-directed"* while students in the traditional curriculum revealed a *"more forward-directed"* type of reasoning.

The CRT was a unidirectional test in which questions had a sequence that did not allow students to freely go back to previous screens to review findings which limits our ability to detect differences between backwards and forward reasoning strategies. However, several research questions were included in the CRT in order to highlight possible differences in reasoning strategies of the students. A first asked the students if they felt they could indicate a differential diagnosis if they needed more information, and was placed after initial history presentations. A second was placed after results of the physical examination and asked the students if those results changed their main differential diagnosis. The analysis of such questions revealed that for the PBL and Integrated curricula CP1 students required more information than their final-year colleagues in order to make a differential diagnosis, while in the traditional curriculum those differences between years (CP1 vs final-year) did not exist. Overall traditional curriculum at both levels seems to require more information to be able to make a differential diagnosis.

Furthermore, it is possible to observe by the results on the second research question (whether or not new findings lead to a change in differential diagnosis), that the majority of PBL and Integrated curricula students (in both years) did not change their differential diagnosis due to the effect of the results of physical examination, while the opposite happened with the students from the traditional curriculum.

One possible explanation may be that traditional curriculum students were reliant on clinical data to be able to identify patterns or generate hypotheses, which could be seen as consistent with Patel *et al.,* (1991) while the others students would be generating hypotheses early on based on history information. But the CRT cases unidirectionality, makes impossible to confirm this.

On the other hand, looking only at the final-year students' answers, where no significant differences in performance were found but differences in strategies seem to emerge. It is possible to see that students in the traditional curricula require more information to indicate a differential diagnosis, are more likely to change an initially presented one on the face of new information, if presented, and tend to indicate their differential diagnosis only after clinical data is presented.

A possible explanation might be that final-year students from the PBL and the Integrated curricula are starting to shift towards a strategy based more on examples,

that being a system 1 pattern recognition model, while traditional curriculum are using a more hypothetico-deductive strategy. That would justify the increased reliance in information used to test previous hypotheses and the more frequent changes in the differential diagnosis. The use of patter recognition by PBL and integrated would justify why they were able to earlier, and with less information, identify differential diagnosis and why those were likely to change on the basis of new information. This can only be understood as an indication, within the limits and particularities of the context of the present research.

It may be that the students are recognising patterns, not from patient examples but from previous scenarios they had encountered in their learning contexts. That would mean these findings would be more a indication of how the students answer diagnostic reasoning 'written' exercises, than actually what diagnostic strategies they would use when faced with clinical problems. Further research should be carried out to investigate this finding and doing so requires further investigation into the validity of the CRT cases (as mentioned previously).

Especially it is important to consider that, the reliance on pattern recognition at this level, even considering these are final-year students, should be looked at with extreme caution and discomfort. Studies have shown patter recognition to be a skill of experts after many years of practice in direct contact with clinical situations (Norman et al. 2007). Although this reasoning strategy allows experienced clinicians to gain efficiency in diagnosis, even when used by these senior clinicians it remains very susceptible to heuristics and bias (Elstein 1999), requiring high developed metacognitive skills to be able to oversee and control this process (Marcum 2012; Croskerry 2009b; Croskerry 2009a). Therefore even final-year students using pattern recognition alone, would be more likely to misdiagnose patients and more importantly less aware that they did so.

B.1 Implication for curriculum development

Here we present the some implications of our study for curriculum development in medical education. We also review the recommendations identified previously in the literature review, in light of our research findings.

Development of clinical reasoning should be an explicit outcome of undergraduate medical curricula.

This recommendation is driven by the analysis of the three curricula considered in the present study, and the national guidelines from the regulatory bodies. In the analysis not many explicit moments and learning opportunities for clinical reasoning development in the early years of medical degree were identified, apart from the PBL curriculum, where some guidance for the PBL cases explicitly focuses on clinical reasoning development and/or differential diagnosis. In all curricula analysed clinical reasoning was an outcome to be achieved throughout clinical placements and clinical environment exposure in the final-years. Nevertheless, in the documentation analysed, this seemed to be a tacit outcome implied in the ability of clinical decision making or differential diagnosis, that the students would develop by the opportunistic contact with clinical situations. There was also only tenuous reference to particular learning/teaching strategies (e.g. shadowing, mentoring) in place to foster the development of such important cognitive ability. Of course, this information is only driven by the analysis of the documents available to the public and the students' handbooks (our sources). Nevertheless, these documents should reflect the ethos of the programmes, as they will be documents the students will rely on to seek information. Therefore, if planned opportunities exist they should be clearly stated, making explicit to the students since the early being of their professional journeys the value and importance of clinical reasoning.

Furthermore clinical reasoning is crucial to the practice of medicine, and it is explicitly mentioned as one of the outcomes defined by the regulatory bodies to be achieved by students upon graduation of medical school. Therefore, it is extremely important that curriculum developers attempt to make the development of this ability an explicit outcome and dedicate efforts to embed explicit opportunities for its development in the different curricula, by creating or making more explicit exiting opportunities and by ensuring assessment strategies used reflect this aim.

Create planned opportunities for clinical reasoning development.

Our analysis of PBL sessions show some indication that these sessions are fostering knowledge integration and maybe fostering some early encapsulation. Additionally, the PBL students and the integrated students at the CP1 level did perform better than their peers in the traditional curriculum. A reason for this may be that the planned integration of knowledge from basic and clinical sciences fostered by these curricula is leading to superior diagnostic ability at this stage.

However, as some have highlighted (A. Harris et al. 2011; Dyrbye et al. 2007; Prince et al. 2000) the PBL cases and teaching vignettes lack the complexity of the real cases, where many emotional, contextual and interpersonal factors are involved (Dyrbye et al. 2007). Also our results show a clear positive effect of exposure to clinical practice for all the curricula involved. Which support the aforementioned idea that the complexity of real clinical context is fundamental for the development of this cognitive ability.

Therefore, in order to help prepare students to become competent doctors able to identify and effectively deal with patients' problems, early opportunities for contact with clinical practice should be provided, and carefully planned in order to ensure the students are developing this critical ability. As Kassirer points *"Rather than using cases artificially constructed from memory, real cases are greatly preferred because they often reflect the false leads, the polymorphisms of actual clinical material, and the misleading test results encountered in everyday practice (…). The teaching of clinical reasoning need not and should not be delayed until students gain a full understanding of anatomy and pathophysiology."* (Kassirer, 2010, p. 1118).

An excellent example of how this is possible is presented by Harris, Boyce and Ajjawi (Harris *et al.* 2011). These authors describe a small group session – clinical reasoning sessions (CRS), in the context of psychiatry. In these sessions a typical PBL case is replaced by a student presenting the findings from the history and physical examination of a real patient done in a clinical placement. This student is asked to look for discriminate features and prioritise patients' problems, the others in the group should "*contribute, suggesting clinical aspects of the case that require further exploration, identifying learning areas that need answers and challenging the diagnostic process employed by the presenter"* (Harris *et al.*, 2011, p. 15). These authors suggest that two sessions should take place with an approximate duration of between 60 to 90 minutes, during which the group and the expert facilitator discuss

different aspects of the cases while simultaneously eliciting the reasoning processes involved (Harris *et al.*, 2011). Although research needs to be done in order to test for positive effects on clinical reasoning development, this is an example of a simple, yet well designed, intervention that could easily be integrated within the various curricula types to create explicit opportunity for clinical reasoning development.

Other suggestions involve case-base conferences, where experts present and discuss real cases with the students while making explicit their own reasoning processes (Jerome P Kassirer 2010); creating opportunities for students in clinical practice to reflect and discuss with mentors/teachers the reasoning strategies that they used (Atkinson et al. 2011); make explicit to the students that in clinical practice they should be seeking to understand similarities between cases even if those seem apparently unrelated (K W Eva 2005a); use 'compare and contrast', forcing the learners to prioritise the differential diagnosis and justify their choices and give the learners feedback (Bowen 2006) and encourage reading around the cases (Bowen 2006).

As previously these are strategies that do not imply major curriculum changes, but to early contact with patients and exposure to clinical environment is critical.

However, it is important to note that although these seem to be aligned with good practices in education, and there are some indications from research reported by these authors that their suggested strategies might foster clinical reasoning. More research is necessary to show that these will have a positive effect on clinical reasoning.

Our results show that the effect of practice was significant overall for both the CRT cases. Nevertheless, when analysing differences at the curricula level, it was possible to observe significant differences between cohorts of students within the traditional and the integrated curriculum, with final-year students achieving better results. However, in the PBL students for CRT case 1, these differences were not significant. For case two, although the differences were significant, they were smaller than those observed in the other two curricula. This seems to show that the effect of practice is dependent on the pre-clinical learning experiences and education environment. A possible explanation is that PBL students are in fact not as well prepared in terms of knowledge and skills to learn from clinical practice, which seems to contradict the results from research into PBL (Albanese & Mitchell, 1993; Schmidt *et al.*, 1996). Another possible explanation is that current clinical placement models might be

more adequate for students from more traditional science-based curricula, and are not prepared to recognise and foster PBL students' previously acquired skills. Each of these possible explanations has great implications for teaching and curriculum development, but more research into this area is needed to have a clear picture. Notwithstanding, our results should be considered: clinical tutors should be aware of the pre-clinical curriculum; and they should know the underlying assumptions and values of that curriculum so they can apply them in teaching in clinical settings. Moreover, medical schools should ensure coherence and alignment between all phases of the curriculum.

Length of clinical placements is very important in planning different types of curriculum.

Our findings suggest that initial differences in diagnostic reasoning ability on the basis of the curricula type, disappear by the time the students finish their degrees. This, of course, needs to be considered in light of the previously discussed limitations of our study. However one aspect is of special interest for curriculum developers, the differences in length of the courses and the clinical placements, that may be the underlining reason for the fading of the differences found in the CP1 cohorts.

The traditional curriculum is a six-year programme, with three full academic years of clinical placements; the integrated curriculum is a five-year degree with two-and-half years of clinical placements and the PBL degree lasts four-and-half years, also with the last two-and-half years of clinical placements. Due to the now recognised importance of clinical practice, it is possible that the extra semester the traditional curriculum students have in clinical practice is contributing towards more even results among the curriculum. This is, of course, a hypothesis and should be understood as such. Nevertheless, from a curriculum development perspective it can be said that the length of clinical placements should be carefully looked at in consideration of the type of curriculum adopted. For example if a school is to develop PBL and traditional-Flexnerian tracks (e.g. an academic MD-PhD) maybe it needs to consider allowing students from the traditional track more time in clinical placements.

A.1. Implications for future research

Achieving a consensual definition of clinical reasoning

This is an implication that arises from the learning journey into clinical reasoning more than specifically from our research findings; however, by its importance it was chosen to include it here. Researchers should make efforts to achieve a consensual definition of clinical reasoning and define nomological networks that would facilitate clinical reasoning research and assessment.

Literature on clinical reasoning is clouded by the lack of a consensual definition. As we highlighted before, this is a common problem of terminology that is used in everyday language and simultaneously in scientific research (Holyoack & Morrison 2005). Simmons did a concept analysis of clinical reasoning in nursing, and highlighted the fact *"Multiple concepts have been used synonymously in the literature: decision-making, problem-solving, clinical judgment, diagnostic reasoning and clinical reasoning."* (Simmons, 2010, p. 1151). The problem being, that although all these concepts are related they are far from being synonymous. This fact makes it extremely difficult to compare studies, critique and juxtapose research findings as although all looking at 'clinical reasoning', in fact the constructs researched differ. Some progress has been made in this direction however, more needs to be done to encourage researchers to agree a consensual definition of this fundamental ability. Another important consequence of achieving such a definition would be that this would be able to inform practice. It is not possible to teach and assess something that we as a research community are still to define what it is. Especially because for those teaching it, clinical teachers and students, also often have their own definitions driven from their own experiences, and these may differ from each other. We need all to be 'speaking the same language'. Achieving this will put the research community a step closer to 'the holy grail' (Schuwirth 2009).

A related aspect is that constructs such as clinical reasoning, that are not observable, rely on nomothetical networks to be researched. These are networks of relationships these concepts establish with others, that are either observable or have on their turn a direct relationship with an observable construct (Trochim 2006). For example, fear is not observable, but research has found that there are behaviours/physical signs that indicate a person is frightened; those are then used to measure fear.

Clinical reasoning research needs to be able to establish these networks, but that will only be possible after a much more basic step has happened: a definition of clinical reasoning is agreed. Psychology and social sciences have a long tradition of doing so, that was the case of so many of the constructs used by these sciences (e.g. personality, verbal reasoning) and certainly medical education and clinical reasoning research can learn valuable lesson from those.

Open the 'black box of PBL and clinical reasoning': challenges ahead

"Clinical decision making is primarily a social and linguistic skill, acquired by participating in communities of practice called health professions. These communities of practice have their own subculture including the language game called clinical decision making which includes an interpretive repertoire of specific language tools and skills." (Loftus, 2006, p. 6)

Taking the quote from Loftus above, if clinical reasoning is a linguistic skill, it will gain from being researched as a linguistic phenomenon, using linguistics methodologies and language interpretative models. Corpus analysis is a possible methodology that allows this type of analysis and is based on sound theoretical grounds within applied linguistics. As it has been shown by Adolphs and colleagues (Adolphs et al. 2005; Adolphs *et al.* 2004) applications to healthcare settings can be extremely useful in highlighting hidden patterns and promoting change.

Regarding its application to clinical reasoning research, our results have shown that it is possible to use this methodology to analyse the verbalised early basis of clinical in case discussions in a PBL setting. However, there is still research to be done in order to be able to ensure this analysis is able to capture such a complex and rich phenomenon as clinical reasoning. A further validation study/studies, similar to the one conducted during this research, but involving large sample of experts, trained to identify aspects such as 'reasoning' utterances and using a larger sample of transcripts should be developed.

Also using corpus analysis to analyse recordings of experts discussions of clinical cases, could provide very useful information as to whether differences in technical language and words per hour, and others found in this study relate with expertise. Also this analysis could provide a useful identification of expertise 'discourse markers' that could then be used to analyse students' discourses.

Additionally based on research using corpus analysis in healthcare (Crawford 1999; Skelton & Hobbs 1999; Skelton *et al.*, 2002; Thomas & A Wilson 1996) we argue this methodology can be applied to the study of clinical reasoning in ward-round discussion between senior clinicians, simulation scenarios, and even doctor-patient recorded interactions, but further research is need to be able to demonstrate this.

Need for longitudinal studies on development of clinical reasoning

Longitudinal studies are needed in order to better understand the development of clinical reasoning. This was one of the major limitations of our research, and of others carried out with the purpose of trying to understand the impact of curricula in the development of clinical reasoning (Goss *et al.* 2011; Arts 2007; Hoffman *et al.* 2006; Schmidt *et al.* 1996; Boshuizen *et al.* 1992). Both the analysis of the PBL session and the study of the differences between different curricula involved comparison between different groups of students, rather than flowing the same group over time.

Longitudinal studies, following groups of students from each different curriculum during their undergraduate medical education, involving multiple data collection moments and instruments, is necessary. The clinical reasoning test developed for the purposes of this research, upon further research into its psychometric characteristics can provide a useful tool to use in such further studies.

Only this type of study can further investigate if in fact, as we found, initial differences exist between curricula and fade with practice, and if so why does this happen? What are the factors driving the learning of students in each of the different curricula? Does in fact, as others have found (Neufeld *et al.* 1981), clinical reasoning ability remain constant through undergraduate medical education.

C. Epilogue

Clinical reasoning has captured the interest, curiosity and fascination of so many researchers, practitioners and educators in the past century. This extraordinary ability to 'solve the biomedical and clinical mysteries', identifying what is the patient's problem and dealing with it effectively is at the heart of modern medicine, and yet, although the good progress of the last decades of research, many questions are yet to be answered. Questions regarding the mental processes involving the weight and importance of knowledge basis, about how individual, situational and contextual factors impact these mental processes, and questions about how we can ensure learning opportunities are provided to students allowing them to develop this ability. As an educationalist and educational researcher by background, I look at clinical reasoning from a development perspective; my interest lies with understanding how students become able to solve clinical problems effectively? What happens during their education and training that is promoting this development? Are there ways in which we, as educationalists can aid these processes? Seeking answers to some of these questions motivated the start of this personal and professional learning adventure, but in the search for answers, new questions have emerged. The current chapter sums up some of the answers found, and presents the new questions.

In the first chapter a *map* of this research journey was presented. Chapter two presented the background of the present research: aspects were reviewed such as a definition of clinical reasoning, its models and types, perspectives on how it develops, and implications for teaching. Chapter two also presented a review of existing assessment instruments and methodologies use to study clinical reasoning.

Chapter three presented the education context and described the curricula studied in the present research.

Chapter four marked the beginning of the amazing journey into words and meanings, and presented the adaptation of a methodology from linguistics to the study of PBL discussions, and described how that methodology was applied in our study. Chapter five presented the results from a study to understand how early stage medical students approach clinical cases. In chapter six the methodology used in a second study conducted as part of this research aimed at understanding the differences and similarities between diagnostic reasoning ability between students at two levels (CP1

and final-year) in three different medical schools. The development and validation of the clinical reasoning assessment tool (CRT) was also presented here. Chapter seven presented the results of the comparison between the curricula of the three medical schools using the CRT, the discussion of those results and the research limitations.

Finally, the present chapter discussed the results of both studies in light of literature in this field, presented the implications for curriculum developments and further research.

The adventure into the mysteries of clinical reasoning started long ago, in a warm place surrounded by inquisitive minds and the generosity of a great mentor. It has been since a rewarding and enriching learning journey. During which I had the pleasure to be surrounded by extraordinary scholars and medical educators, role models. Their inquisitive questioning and critical views were fundamental to the present work.

It is my belief that a PhD should be much more than just learning 'a lot about very little', it should an opportunity to learn, develop as a scholar, further develop as a researcher, employing skills such as scientific inquire, critical thinking and analysis to understand reality/realities and contribute towards the advance of scientific knowledge. However, I have come to realise while embarking in this journey, is that this is only the start of an immense learning journey and of what, I hope, will be a very fruitful contribution to knowledge and research in medical education. The work here presented is my small contribution to this body of knowledge, moreover is a new beginning that could not be better described than by the Flexner's quote used in the beginning of this chapter describes, as new questions simmer through my mind as I write these words.

Every great advance in science has issued from a new audacity of imagination.

John Dewey

Bibliography:

Adolphs, S *et al.*, 2004. Applying corpus linguistics in a health care context. *Journal of Applied Linguistics, Vol.*, 1(1), pp.9–28.

Adolphs, Svenja et al., 2005. Clinical Linguistics? Corpus Linguistics in Health Care Settings. *Language.[online]*. Available at: http://www.brown.uk.com/teaching/city/datadriven.pdf.

Adolphs, Svenja, 2006. *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*, New York, U.S.A.: Routledge.

Adolphs, Svenja & Carter, Ronald, 2007. Beyond the word. *European Journal of English Studies*, 11(2), pp.133–146.

Ajjawi, R. & Higgs, J, 2008. Learning to reason: a journey of professional socialisation. *Advances in health sciences education : theory and practice*, 13(2), pp.133–50.

Albanese, M. A. & Mitchell, S. 1993. Problem-based learning: a review of literature on its outcomes and implementation issues. Academic Medicine. 68 (1), 52–81.

Albanese, M. A. 2000. Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. Medical education. 34 (9), 729–738.

Allen, M.D. & Fong, A.K., 2008. Clinical application of standardized cognitive assessment using fMRI. I. Matrix reasoning. *Behavioural Neurology*, 20(3), pp.127–140.

Alexander, H.A. 2006. A View from Somewhere: Explaining the Paradigms of Educational Research. Journal of Philosophy of Education. 40 (2), 205–221.

Anderson, K. et al., 2008. The assessment of student reasoning in the context of a clinically oriented PBL program. *Medical teacher*, 30(8), pp.1–8.

Anderson, K.J., 2006. *Factors Affecting The Development Of Undergraduate Medical Students' Clinical Reasoning Ability*. Ph.D. Thesis: University Of Adelaide.

Archer, D., Wilson, Andrew & Rayson, Paul, 2002. *INTRODUCTION TO THE USAS CATEGORY SYSTEM*, Lancaster: UCREL, University Centre for Computer Corpus Research on Language. Available at: http://www.comp.lancs.ac.uk/ucrel/usas/.[Accessed December 8, 2010].

Ark, T.K., Brooks, Lee R & Eva, Kevin W, 2006. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Academic medicine : journal of the Association of American Medical Colleges*, 81(4), pp.405–9.

Arppe, A. et al., 2010. Cognitive Corpus Linguistics : five points of debate on current theory and methodology. *Communication*, 5(1), pp.1–27.

Arts, J., 2007. Developing managerial expertise: studies on managerial cognition and the implications for management education.

Arts, J., Gijselaers, Wim H. & Segers, M.S.R., 2006. Enhancing problem-solving expertise by means of an authentic, collaborative, computer supported and problem-based course. *European Journal of Psychology of Education*, 21(1), pp.71–90.

Arts, J, Gijselaers, W.H. & Boshuizen, H., 2000. Expertise Development in Managerial Sciences: The Use of Knowledge Types in Problem-Solving. In *Annual Meeting of the American Educational Research Association*. New Orleans, LA, p. 21.

Atkinson, K., Ajjawi, R. & Cooling, N., 2011. Promoting clinical reasoning in general practice trainees: role of the clinical teacher. *The clinical teacher*, 8(3), pp.176–80.

Baker, A.C., Jensen, P.J. & Kolb, David A., 2002. *Conversational learning: an experiential approach to knowledge creation*. Greenwood: Greenwood Publishing Group.

Baker, C.M., 2000. Problem-based learning for nursing: integrating lessons from other disciplines with nursing experiences. *Journal of professional nursing : official journal of the American Association of Colleges of Nursing*, 16(5), pp.258–66.

Barrows, H. S., 1986. A taxonomy of problem-based learning methods. *Medical Education*. 20(6), pp.481-486.

Barrows, H S & Feltovich, P J, 1987. The clinical reasoning process. *Medical Education,* 21(2), pp.86–91.

Barrows, Howard S., 1996. Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 1996(68), pp.3–12.

Barrows, H S, Evensen, D.H. & Hmelo, C E, 2000. Foreword. In Mahwah: Lawrence Erlbaum, pp. vii–ix.

Ber, R., 2003. The CIP (comprehensive integrative puzzle) assessment method. *Medical teacher*, 25(2), pp.171–6.

Bereday, G.Z.F., 1964. *Comparative method in education*.Holt, Rinehart and Winston.

Bereiter, C. & Scardamalia, M., 1999. Process and product in PBL research. *Toronto: Ontario Institutes for Studies in Education/University of Toronto.*

Berkel, H. & Schmidt, Henk G, 2000. Motivation to commit oneself as a determinant of achievement in problem-based learning. *Higher Education*, 40, pp.231–242.

Berliner, D.C., 2002. Educational Research:The Hardest Science of All. *Educational Researcher*, 31(8), pp.18–20.

Berner *et al.*, 1974. Evaluation Problem-Solving. Journal of Medical Education. 49666–672.

Berner, E.S., 2011. Mind wandering and medical errors. *Medical education*, 45(11), pp.1068–9.

Berry, D.C. & Broadbent, D.E., 1984. On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology Section A*, 36(2), pp.209–231.

Beullens, J. *et al.*2005. Do extended matching multiple-choice questions measure clinical reasoning? *Medical education.* 39 (4), 410–417.

Biesta, G.J.J. & Burbules, N.C., 2003. *Pragmatism and educational research*. Lanham , MD: Rowman & Littlefield.

Bordage, G & Zacks, R., 1984. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Medical Education*, 18(6), pp.406–416.

Bordage, G, 1987. The curriculum: overloaded and too general? *Medical Education.* 21(3), pp.183–188.

Bordage, G, Grant, J. & Marsden, P., 1990. Quantitative assessment of diagnostic ability. *Medical Education*, 24, pp.413–425.

Bordage, G & Lemieux, M., 1991. Semantic structures and diagnostic thinking of experts and novices. , 66(9 Suppl), pp.S70–2.

Bordage, G, 1994. Elaborated knowledge: A Key to Sucessful Diagnostic Thinking. *Academic Medicine*, 69(11), pp.883–5.

Bordage, G et al., 1995. Content validation of key features on a national examination of clinical decision-making skills. , 70(4), pp.276–281.

Bordage, G, 1999. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine*, 74(10).

Bordage, G et al., 1997. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. *Academic Medicine;*72(10 Suppl 1), pp.S37–9.

Bordage, Georges, 2007. Moving the field forward: going beyond quantitative-qualitative. *Academic medicine : journal of the Association of American Medical Colleges*, 82(10 Suppl), pp.S126–8.

Bordage, G.s, 2007b. Prototypes and semantic qualifiers: from past to present. *Medical education*, 41(12), pp.1117–21.

Borsboom, D., 2005. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*, Cambridge, UK: Cambridge University Press.

Borsboom, D., 2008. Psychometric Perspectives on Diagnostic Systems. *Online*, 64(9), pp.1089–1108.

Boshuizen, H. *et al.,* 1992. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science: A Multidisciplinary Journal*, 16(2), pp.153–184.

Boshuizen, H et al., 1995. Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction*, 5(4), pp.269–289.

Boshuizen, H.P A et al., 1997. Measuring knowledge and clinical reasoning skills in a problem-based curriculum. *Medical Education*, 31(2), pp.115–121.

Boshuizen, H.P.A. & Schmidt, H.G., 2000. *The development of clinical reasoning expertise*. In: Higgs J, Jones M(eds) Clinical reasoning in the health Professions 2nd Edn. Oxford, UK: Butterworth-Heinemann, p.15-22

Boshuizen, H.P.A., 2003. Expertise development : How to bridge the gap between school and work. In H P.A. Boshuizen, ed. *Expertise development: The transition between school and work*. Open Universiteit Nederland, pp. 7–31.

Van den Bossche, P, Gijbels, D & Dochy, F, 2000. Does problem based learning educate problem solvers? A Meta-analysis on the effects of problem based learning.

Botezatu, M., Hult, H. & Fors, U.G., 2010. Virtual patient simulation: what do students make of it? A focus group study. *BMC medical education*, 10(1), p.91.

Bowen, J.L., 2006. Educational strategies to promote clinical diagnostic reasoning. *The New England journal of medicine*, 355(21), pp.2217–25.

Boulouffe, C. *et al.* 2010. Evaluation of Clinical Reasoning in Basic Emergencies Using a Script Concordance Test. American Journal of Pharmaceutical Education. 74 (10), 194.

Bray, M., 2005. Methodology and focus in comparative education. In *Education and Society in Hong Kong and Macao*. pp. 237–251. Available at: http://www.springerlink.com/index/KX8T6147609111J4.pdf [Accessed December 21, 2012].

Bray, M., Adamson, B. & Mason, M.(eds), 2007. *Comparative Education Research: Approaches and Methods*. Hong-Kong: CERC Springer.

Bland AC, Kreiter CD, Gordon JA. 2005. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med* 80:pp.395–399.

Broadfoot, P., 2010. Comparative Education for the 21st Century : retrospect and prospect. *Comparative Education*, 36:3(23), pp.37–41.

Brooks, L R, Norman, G R & Allen, S.W., 1991. Role of specific similarity in a medical diagnostic task. *Journal of experimental psychology. General*, 120(3), pp.278–87.

Brown, Brian, Crawford, Paul & Carter, Ronald, 2006. *Evidence-based health communication*, Maidenhead: Open University Pres, McGraw-Hill.

Bunniss, S. & Kelly, D.R. 2010. Research paradigms in medical education research. *Medical Education*: 44 (4), 358–366.

Cameron, R., 2011. Mixed Methods Research: The Five Ps Framework. *The Electronic Journal of Business Research Methods Volume*, 9(2), pp.96–108.

Carey, J.W. et al., 1995. Intercoder Agreement in Analysis of Responses to Open-Ended Interview Questions : Examples from Tuberculosis Research. *Cultural Anthropology methods*, 8(3), pp.1–5.

Caire F,*et al.,*2004: Autoévaluation des internes en neurochirurgiepar tests de concordance de script (TCS): le processus d'élaboration des tests. [Self- assessment for neurosurgery residents by script concord- ance test (SCT). The process of test elaboration]. *Neurochirurgie*, 50:pp.66-72.

Carrol, L., 1871. *Through the Looking-Glass* Ebook edit., London, England: The Project Gutenberg.. Available at: http://www.gutenberg.org/files/12/12-h/12-h.htm. [Accessed: 10 December 2011]

Carver, S.M. & Klahr, D., 2001. *Cognition and instruction: Twenty-five years of progress.* Mahwah NJ:Lawrence Erlbaum Associates.

Castel, A.D. *et al.*, 2007. The dark side of expertise: domain-specific memory errors. *Psychological Science*, 18(1), pp.3–5.

Charlin, B, Mann, K. & Hansen, P., 1998. The many faces of problem-based learning: a framework for understanding and comparison. *Medical Teacher*, 20(4), pp.323–330.

Charlin, B, *et al.*, 2000. The Script Concordance test: a tool to assess the reflective clinician. *Teaching and Learning in Medicine*, 12(4), pp.189–195.

Charlin, B, Tardif, J. & Boshuizen, H.P., 2000b. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic medicine : journal of the Association of American Medical Colleges*, 75(2), pp.182–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10693854 [Accessed November 4, 2010].

Charlin, B., *et al.*, 2007. Scripts and clinical reasoning. *Medical education*, 41(12), pp.1178–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18045370 [Accessed July 6, 2011].

Chase, G. & Simon, H. A, 1973. Perception in Chess. *Public Health*, 4(1), pp.55–61.

Cobb, P. et al., 2003. Design Experiments in Educational Research. *Educational Researcher*, 32(1), pp.9–13.

Cohen LJ, Fitzgerald SG, Lane S, Boninger ML. 2005 Development of the seating and mobility script con- cordance test for spinal cord injury: obtaining content validity evidence. *Assist Technol.*;17pp.122–32.

Colliver, J.A et al., 1990. Case Specificity of Standardized-Patient Examinations: Consistency of Performance on Components of Clinical Competence Within and Between Cases. , 13(2), pp.252–261.

Colliver, J., 2000. Effectiveness of problem-based learning curricula: research and theory. *Academic medicine : journal of the Association of American Medical Colleges*, 75(3), pp.259–66.

Conrad, C.F., 2006. *The SAGE Handbook for Research in Education: Engaging Ideas and Enriching Inquiry*. London:Sage Publications, Inc.

Cook, C., Heath, F. & Thompson, R.L., 2000. A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, 60(6), pp.821–836.

Cook, D.A. & Triola, M.M., 2009. Virtual patients: a critical literature review and proposed next steps. *Medical education*, 43(4), pp.303–311.

Cooke, M. *et al.,* 2007. American medical education 100 years after the Flexner report. *Archives of Iranian medicine*, 10(1), pp.131–133.

Cooksey, R., 2000. Commentary on Cognitive adaptation and its consequences: a test of cognitive continuum theory. *Journal of Behavioral Decision Making*, 13(1), pp.35– 54.

Coughlin, L. & Patel, V L, 1985. Adapting a paradigm from cognitive science to medical education: problems and possible solutions. , 24, pp.97–102.

Coughlin, L.D. & Patel, V.L., 1987. Processing of Critical information by Physicians and Medical students. *Academic Medicine*, 62(10), p.818.

Crawford, P, 1999. A language for nursing: furthering the debate. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, 14(12), pp.36–7.

Crawford, Paul, Brown, Brian & Nerlich, B., 2008. The “moral careers” of microbes and the rise of the matrons: An analysis of UK national press coverage of methicillin-resistant Staphylococcus aureus (MRSA) 1995–2006. *Health, Risk & Society*, 10(4), pp.331–347.

Crawford, P *et al.,* 2010. The Language of compassion. *Taiwan International ESP Journal*, 3(1), pp.1–16.

Crawford, P & Brown, B.J., 2010. Health communication: Corpus linguistics, data driven learning and education for health professionals. *International English for Specific Purposes Journal*, 2, pp.1–26.

Creswell et al., 2003. Advanced mixed methods research designs. In Tashakkori & Teddlie, eds. *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, California: SAGE, pp. 209–240.

Creswell, J.W. & Clark, V.L.P., 2007. Understanding mixed methods research. In *Designing and conducting mixed methods research.*(eds) Thousand Oaks, California: Sage Publications, Inc, pp. 1–19.

Creswell, J.W., 2009. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, SAGE Publications.

Creswell, J.W. & Clark, V.L.P., 2011. Foundations of mixed methods research. In *Designing and Conducting Mixed Methods Research*. Sage Publications, Inc.

Croskerry, P., 2009a. A universal model of diagnostic reasoning. *Academic medicine : journal of the Association of American Medical Colleges*, 84(8), pp.1022–8.

Croskerry, P., 2009b. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education*, 14(0), pp.27–35.

Crossley, M., 2010. Bridging Cultures and Traditions Comparative and International Education Bridging Cultures and Traditions in the Reconceptualisation of Comparative and International Education. , (December 2012), pp.37–41.

Custers, E J, Boshuizen, H.P. & Schmidt, H G, 1996a. The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. , 24(3), pp.384–399.

Custers, E., Regehr, G. & Norman, G. R., 1996b. Mental Representations of Medical Diagnostic Knowledge: A Review. *Acad Med*, 71(10), pp.S55–S61.

Dale, R., 2005. Globalisation, knowledge economy and comparative education. *Comparative Education*, 41(2), pp.117–149.

DeGroot, 1946. *Through and Choice in Chess [Original work]*, The Hague: Mouton.

Dennick, R, 2008. Theories of learning: Constructive experience. In D. Matheson, ed. *An introduction to the study of education*. London, England: Routledge London, p. 29.

Dent, J.A. & Harden, Ronald M, 2005. *A practical guide for medical teachers*, University of Michigan: Elsevier Churchill Livingstone.

Dochy, Filip *et al.*, 2003. Effects of problem-based learning: a meta-analysis. , 13(5), pp.533–568.

Dory, V., Gagnon, R. & Charlin, B, 2009. Is case-specificity content-specificity? An analysis of data from extended-matching questions. *Adv Health Sci Educ Theory Pract.* 15(1), pp.55-63.

Duffy, T.P., 2011. The Flexner Report--100 years later. *The Yale journal of biology and medicine*, 84(3), pp.269–76.

Durning, S. *et al.,* 2011. Context and clinical reasoning: understanding the perspective of the expert's voice. *Medical education*, 45(9), pp.927–38.

Durrett, R., 2010. *Probability: Theory and Examples* 4, illustr., Cambridge University Press.

Dyrbye, L.N., Harris, I. & Rohren, C.H., 2007. Early clinical experiences from students' perspectives: a qualitative study of narratives. *Academic medicine : journal of the Association of American Medical Colleges*, 82(10), pp.979–88.

Elstein, A., Shulman, L S & Sprafka, S A, 1978. *Medical Problem Solving: an Analysis of Clinical Reasoning*, Cambridge, MA: Harvard Un. Press.

Elstein, A S, Shulman, L S & Sprafka, S A, 1981. Medical problem-solving. *Academic medicine : journal of the Association of American Medical Colleges*, 56(1), pp.75–76.

Elstein, A S & Holmes, M.M., 1981. The limits of rational decision making: anthropological and psychological perspectives. , 5(4), pp.340–344.

Elstein, Arthur S, Shulman, Lee S & Sprafka, Sarah A, 1990. Medical Problem Solving: A Ten-Year Retrospective. *Eval Health Prof*:13(1), pp.5–36.

Elstein, Arthur S & Bordage, G., 1991. Phychology of Clinical Reasoning. In J. Dowie & A S Elstein, (eds). *Professional judgment: a reader in clinical decision making*. Cambridge: Cambridge University Press, pp. 109–129.

Elstein, A S, 1999. Heuristics and biases: selected errors in clinical reasoning. , 74(7), pp.791–794.

Elstein, Arthur S & Schwartz, A., 2002. Evidence base of clinical diagnosis: Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*, 324(7339), pp.729–732.

Elstein, A S, 2009. Thinking about diagnostic thinking: a 30-year perspective. , 14 Suppl 1, pp.7–18.

Engel, C.E., Boud, D. & Felletti, G.P., 1991. Not Just a Method but a Way of Learning. In D. Boud & GP Feletti, (eds). *The Challenge of Problem-Based Learning*. London: Kogan Page, pp. 22–33.

Epstein, R M, 1999. Mindful practice. *JAMA : the journal of the American Medical Association*, 282(9), pp.833–839.

Epstein, Ronald M., 2007. Assessment in medical education. *The New England journal of medicine*, 356(20), p.2109;

Ericsson, K Anders & Simon, Herbert A, 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: The MIT Press.

Ericsson, K.A., Krampe, R.T. & Tesch-Römer, C., 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), p.363.

Ericsson, K. & Simon, H A, 1998. How to Study Thinking in everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, 5(3), pp.178–186.

Ericsson, K A, 2008. Protocol Analysis and Verbal Reports on Thinking [webpaper]. Available at: http://www.psy.fsu.edu/faculty/ericsson/ericsson.proto.thnk.html [Accessed December 10 2010]

Ericsson, K A & Simons, H.A., 1993. *Protocol Analysis - Verbal Reports As Data (Revised Edtion)* , Massachusetts, U.S.A.: MIT Press.

Ericsson, KA Anders, 2004. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Acad. Med.*, 79(10), pp.70–81.

Ericsson, K Anders, 2006a. Protocol Analysis and Expert Thought : Concurrent Verbalizations of Thinking during Experts ' Performance on Representative Tasks. In Ericsson, K.A (eds). *Cambridge Handbook of Expertise*. Cambridge MA: Cambridge University Press.

Ericsson, K Anders, 2007. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical education*, 41(12), pp.1124–30.

Ericsson, K. Anders & Ward, P., 2007. Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance. *Current Directions in Psychological Science*, 16(6), pp.346–350.

Ericsson, K. Anders & Towne, T.J., 2010. Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(3), pp.404–416.

Eva, K W, Neville, A J & Norman, G R, 1998. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Academic Medicine*;73(10 Suppl), pp.S1–5.

Eva, K.W. et al., 2002. Expert-Novice Differences in Memory: A Reformulation. *Teaching and Learning in Medicine*, 14(4), pp.257–263.

Eva, K W, 2003. On the generality of specificity. *Med Educ.* , 37(7), pp.587–588.

Eva, Kevin W *et al.,* 2004. How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in health sciences education theory and practice*, 9(3), pp.211–224.

Eva, K W, 2005. What every teacher needs to know about clinical reasoning. *Med Educ.*, 39(1), pp.98–106.

Eva, Kevin W & Norman, Geoffrey R, 2005. Heuristics and biases-a biased perspective on clinical reasoning. *Medical education*, 39(9), pp.870–2.

Eva, Kevin W & Lingard, L., 2008. What's next? A guiding question for educators engaged in educational research. *Medical education*, 42(8), pp.752–4.

Evans, J.S. et al., 2000. Frequency versus probability formats in statistical word problems. *Cognition*, 77(3), pp.197–213.

Evans, J.S.B.T., 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual review of psychology*, 59, pp.255–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18154502.

Evans, S. & Keenan, T., 2009. *An introduction to child development* 2nd Editio., London, England: Sage publications.

Facione, P.A., 2010. Critical Thinking : What It Is and Why It Counts. *Thinking*, (2008), pp.1–24.

Feilzer, M.Y., 2009. Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research*, 4(1), pp.6–16.

Feldon, D.F., 2006. The Implications of Research on Expertise for Curriculum and Pedagogy. *Educational Psychology Review*, 19(2), pp.91–110.

Feltovich, P.J. & Barrows, H.S, 1984. Issues of generality in medical problem solving. In H.G. Schmidt & M. L. DeVolder, eds. *Tutorials in Problem-Based Learning*. Assen, the Netherlands: Van Gorcum, pp. 128–142.

Feltovich, P., Prietula, M. & Ericsson, K., 2006. Studies of expertise from psychological perspectives. In Ericsson, K.A (eds). *Cambridge Handbook of Expertise*. Cambridge MA: Cambridge University Press.

FMUC.[no date] Faculdade de Medicine Universidade de Coimbra: Mapa de Conteudos [online]. Available from: http://www.uc.pt/fmuc/ensino/mim/contentorDocumentos/MapaOp [Accessed 19 October 2011].

FMUC.[no date]b Faculdade de Medicine Universidade de Coimbra: Plano Curricular Mestrado Integrado Medicina.[online] Available from: https://intra.fmed.uc.pt/[Accessed 19 October 2011].

FMUC [no date] Universidade de Coimbra: Pagina da Faculdade de Medicina [online] Available from: http://www.uc.pt/fmuc[Accessed 19 May 2011].

Friedman, D.S. et al., 2009. Doctor–Patient Communication in Glaucoma Care: Analysis of Videotaped Encounters in Community-Based Office Practice. *Ophthalmology*, 116(12), pp.2277–2285.e3.

Garrud, P., 2009. Case 9.01 Hard labour. University of Nottingham:Gradute Entry Medical School; pp.1–22.

Garrud, P., 2011. Who applies and who gets admitted to UK graduate entry medicine? - an analysis of UK admission statistics. *BMC medical education*, 11, p.71.

Van Gessel, E. *et al.*, 2003. Development of clinical reasoning from the basic sciences to the clerkships: a longitudinal assessment of medical students' needs and self-perception after a transitional learning unit. *Medical education*, 37(11), pp.966–974.

Gigerenzer, G. & Hoffrage, U., 1995. How to Improve Bayesian Reasoning Without Instruction : Frequency Formats. *Psychological Review*, (4), pp.684–704.

Gigerenzer, G. & Hoffrage, U., 1999. Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, 106(2), pp.425–430.

GAMSATUK, 2011. Graduate Medical School Admissions Test. Available at: http://www.gamsatuk.org/index.php?option=com_content&view=article&id=12&Itemid=48 [Accessed 08 October 2011]

GMC, 2009. GEneral Medical Council: *Tomorrow 's Doctors* [online]. Available at: http://www.gmc-uk.org/education/undergraduate/undergraduate_policy/tomorrows_doctors.asp. [Accessed 12 November 2012].

Gobet, F., 2007. A pattern-recognition theory of search in expert problem solving. *Thinking Reasoning*, 3(4), pp.291–313.

Godden, D. & Baddeley, A., 1975. Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8295.1975.tb01468.x/abstract [Accessed January 13, 2012].

Golafshani, N., 2003. Understanding Reliability and Validity in Qualitative Research. *Test*, 8(4), pp.597–606.

Goldkuhl, G., 2008. Practical inquiry as action research and beyond. *European Conference on Information Systems, 2008*, (2004). Available at: http://swepub.kb.se/bib/swepub:oai:DiVA.org:liu-44169 [Accessed March 3, 2013].

Gooberman-Hill, R. *et al.,* 2009. Expressing pain and fatigue: a new method of analysis to explore differences in osteoarthritis experience. *Arthritis and rheumatism*, 61(3), pp.353–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19248131 [Accessed March 16, 2013].

Goss, B. et al., 2011. Comparison of medical students' diagnostic reasoning skills in a traditional and a problem based learning curriculum. *International Journal of Medical Education*, 2, pp.87–93.

Goulet, F. *et al.* 2010. Poorly performing physicians: does the Script Concordance Test detect bad clinical reasoning? *The Journal of continuing education in the health professions:* 30 (3).pp.161–166.

Graber, M.L., 2009. Educational strategies to reduce diagnostic error: can you teach this stuff? *Advances in Health Sciences Education*, 14(0), pp.63–69.

Groen, G. J. & Patel, Vimla L., 1985. Medical problem-solving: some questionable assumptions. *Medical Education*, 19(2), pp.95–100.

Groves, M., 2002. *THE CLINICAL REASONING PROCESS: A STUDY OF ITS DEVELOPMENT IN MEDICAL STUDENTS*. Ph.D Thesis:University of Queensland.

Groves, M., Scott, I. & Alexander, Heather, 2002. Assessing clinical reasoning: a method to monitor its development in a PBL curriculum. *Medical teacher*, 24(5), pp.507–15.

Groves, M, O'Rourke, P. & Alexander, H, 2003. The association between student characteristics and the development of clinical reasoning in a graduate-entry, PBL medical programme. , 25(6), pp.626–631.

Guenter, D., Fowler, N. & Lee, L., 2011. Clinical uncertainty: Helping our learners. *Canadian family physician Médecin de famille canadien*, 57(1), pp.120–2.

Gwet, K., 2002. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. *Statistical methods for inter-rater reliability*, (2), pp.1–9.

Hagan, Pam, 2009. Case 9.02 Blood Brothers. University of Nottingham: Graduate Entry Medical School , pp.1–15.

Hak, T. & Maguire, P., 2000. Group process: the black box of studies on problem-based learning. *Academic Medicine*, 75(7), pp.769–72.

Harden, R., Sowden, S. & Dunn, D., 1984. ASME Medical Education Booklet No. 18. *ASME GUIDE*.

Harden, R.M. *et al*., 1998. The continuum of problem-based learning. *Medical Teacher*, 20(4), pp.317–322.

Harden, R M, 2000. The integration ladder: a tool for curriculum planning and evaluation. *Medical Education*, 34(7), pp.551–7.

Harris, A., Boyce, P. & Ajjawi, R., 2011. Clinical reasoning sessions: back to the patient. *The clinical teacher*, 8(1), pp.13–6.

Harvey, K.J. et al., 2007. "Am I normal?" Teenagers, sexual health and the internet. *Social science & medicine* , 65(4), pp.771–81.

Harvey, K. et al., 2008. Health communication and adolescents: what do their emails tell us? *Family practice*, 25(4), pp.304–11.

Harvey, K., 2012. Disclosures of depression: Using corpus linguistics methods to examine young people's online health concerns. *International Journal of Corpus Linguistics*, 17(3), pp.349–379.

Hayes, S.C., 1986. The case of the silent dog-Verbal reports and the analysis of rules: A review of Ericsson and Simon's protocol analysis:Verbal reports as data. *Journal of experimental analysis of behavior*, 3(3), pp.351–363.

Heemskerk, L. *et al.,* 2008. The effect of question format and task difficulty on reasoning strategies and diagnostic performance in Internal Medicine residents. , 13(4), pp.453–462. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17237966.

Hewett, D.G. *et al.,* 2009. Intergroup communication between hospital doctors: implications for quality of patient care. *Social science & medicine (1982)*, 69(12), pp.1732–40.

Heyvaert, M., Maes, B. & Onghena, P., 2011. Mixed methods research synthesis: definition, framework, and potential. *Quality & Quantity*, 47(2), pp.659–676.

Higgs, Joy & Jones, Mark A., 2000. *Clinical Reasoning in Health Professions* 2nd ed. Higgs &Jones (eds)., Elsevier Health Sciences.

Higgs, J, Jones, M A & Christensen, N., 2008. *Clinical Reasoning in Health Professions* 3rd HIggs, *et al* (eds), Elsevier:Butterworth Heinemann.

Hmelo-Silver, Cindy E, 1998. Cognitive Consequences of Problem-Based Learning for the Early Development of Medical Expertise. *Teaching and Learning in Medicine*, 10(2), pp.92–100.

Hmelo-Silver, C.E. & Barrows, Howard S, 2008. Facilitating Collaborative Knowledge Building. *Cognition and Instruction*, Feb, pp.37–41.

Hmelo-Silver, C.E., Chernobilsky, E. & Jordan, R., 2008. Understanding collaborative learning processes in new learning environments. *Instructional Science*, 36(5-6), pp.409–430.

Hobus, P.P. et al., 1987. Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. , 21(6), pp.471–476.

Hoffman, K. *et al*., 2006. Problem-based learning outcomes: ten years of experience at the University of Missouri-Columbia School of Medicine. , 81(7), pp.617–625.

Holyoack, K.J. & Morrison, R.G., 2005. *The Cambridge Handbook Of Thinking and Reasoning* , New York, U.S.A.: Cambridge University Press.

Hubbard, J. P. et al. (1965) An Objective Evaluation of Clinical Competence. New Technics Used by the National Board of Medical Examiners. 2721321–1328. [online]. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14299142. [Accessed 10 Februray 2009].

Hughes, G. et al., 2008. Case 2.03 It's affecting my football. *University of Nottingham: Graduate Entry Medical School,* pp.1–18.

Hughes, G., 2008a. Case 2.04 Can we fix it ? *University of Nottingham: Graduate Entry Medical School,*, pp.1–18.

Hughes, G., 2008b. Case 8.06 A Shaky Serve. *University of Nottingham: Graduate Entry Medical School*.pp.1-19.

Irving, W., 2009. Case 9.03 More than just a rash. *University of Nottingham: Graduate Entry Medical School,*pp.1–19.

Jarvis, P., 2010. Education Globalisation , the Learning Society and Comparative Education. , (December 2012), pp.37–41.

Jennings, P. & Dhindsa, P., 2009. Case 9.04 Put your foot down. , pp.1–18.

Jiva, D. & Teja, B., 2012. The Medical School Application Guide UK Undergraduate Applicants.

Jones, R. et al., 2001. Changing face of medical curricula. *The Lancet*, 357, pp.699–703. Available at: http://www.sciencedirect.com/science/article/pii/S0140673600041349 [Accessed November 21, 2012].

Jong, T., 2009. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38(2), pp.105–134. Available at:

http://www.springerlink.com/index/10.1007/s11251-009-9110-0 [Accessed July 31, 2011].

Kaplowitz, M.D., Hadlock, T.D. & Levine, R., 2004. A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, 68(1), pp.94–101. Available at: http://poq.oxfordjournals.org/cgi/doi/10.1093/poq/nfh006.

Kassirer, J P & Kopelman, R I, 1989. Learning clinical reasoning from examples. *Hospital practice Office ed*, 24(3), pp.27, 32–34, 44–45.

Kassirer, Jerome P, Wong, J.B. & Kopelman, Richard I., 2009. *Learning Clinical Reasoning* 2nd Revise., Philadelphia/US: Lippincott Williams and Wilkins.

Kassirer, Jerome P, 2010. Teaching Clinical Reasoning : Case-Based and Coached. *Academic Medicine*, 85(7), pp.1118–1124.

Kieser, J.A., Dall'alba, G. & Livingstone, V., 2009. Impact of curriculum on understanding of professional practice: a longitudinal study of students commencing dental education. , 14(3), pp.303–314.

Kintsch, W., 2000. Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review*. 7(2),pp. 257-266.

Koh, G.C. et al., 2008a. The effects of problem-based learning during medical school on physician competency: a systematic review. *CMAJ*, 178(1), pp.34–41.

Koh, G.C. et al., 2008b. The effects of problem-based learning during medical school on physician competency: a systematic review. , 178(1), pp.34–41.

Kolb, D.A., 1984. *Experiential learning : experience as the source of learning and development*, Englewood Cliffs: NJ: Prentice Hall.

Kolb, D.A., Boyatzis, R.E. & Mainemelis, C., 2000. Experiential learning theory: Previous research and new directions. In R. J. Sternberg & L. F. Zhang, eds. *Perspectives on thinking, learning, and cognitive styles*. Lawrence Erlbaum, pp. 227–247.

Kolb, D.A. & Kolb, A.Y., 2005. Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4(2), pp.193–212..

Koller, V. et al., 2008. Using a semantic annotation tool for the analysis of metaphor in discourse. *metaphorik.de*, 15, pp.141–160.

Kondracki, N.L., Wellman, N.S. & Amundson, D.R., 2002. Content analysis: review of methods and their applications in nutrition education. *Journal of nutrition education and behavior*, 34(4), pp.224–30.

Koschmann, T., Evensen, D.H. & Hmelo, C E, 2000. Five readings of a Single Text: Transcripts of a Video analysis Session. In London: Lawrence Erlbaum Associates, pp. 137–166.

Koschmann, T. & MacWhinney, B., 2001. Opening the black box: why we need a PBL talkbank database. , 13(3), pp.145–147.

Koteyko, N., Brown, Brian & Crawford, Paul, 2008. The Dead Parrot and the Dying Swan: The Role of Metaphor Scenarios in UK Press Coverage of Avian Flu in the UK in 2005–2006. *Metaphor and Symbol*, 23(4), pp.242–261.

Kravchenko, A. V, 2002. Cognitive Linguistics As A Methodological Paradigm. In *Cognitive Linguistics Today*. Frankfurt/Main: Peter Lang, Pp. 41–55.

Kuiper, R. et al., 2008. Debriefing with the OPT model of clinical reasoning during high fidelity patient simulation. *Int J Nurs Educ Scholarsh*., 5, p17.

Kuiper, R.A. & Pesut, D.J., 2004. Promoting cognitive and metacognitive reflective reasoning skills in nursing practice: Self-regulated learning theory. *Journal of Advanced Nursing*, 45(4), pp.381–391.

Kuipers, B., Moskowitz, A.J. & Kassirer, J P, 1988. Critical decisions under uncertainty: representation and structure. *Cognitive science*, 12, pp.177–210.

Kurzenhäuser, S. & Hoffrage, U., 2002. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Medical teacher*, 24(5), pp.516–21.

Kwak, N. & Radler, B., 2002. A Comparison Between Mail and Web Surveys : Response Pattern , Respondent Profile and Data Quality. *Journal of Official Statistics 1*8(2) pp.257-73.

Landauer & Laham, 1998. Introduction to latent semantic analysis. *Discourse processes*, 25, pp.259–284.

Lane, J. & Slavin, S., 2001. Simulation in medical education: A review. *Simulation Gamin.* 32: pp.297

Larkin, J. et al., 1980. EXPERT and NOVICE PERFORMANCE in SOLVING PHYSICS PROBLEMS. *Science*, 208(June), p.1980.

Laxmisan, A. et al., 2005. Decisions about critical events in device-related scenarios as a function of expertise. , 38(3), pp.200–212.

Lebeau, J. & Pagonis, D. (2006) Script concordance test: how to evaluate medical reasoning in uncertain situations. *Revue de stomatologie et de chirurgie maxillofacial.*107 (5) pp.327–329.

Leblanc, V R, Norman, G R & Brooks, L R, 2001. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. , 76(10 Suppl), pp.S18–20.

Leblanc, Vicki R, Bould, M.D. & Sharma, B., 2011. 18 Simulation in Postgraduate Medical Education. *Simulation*.

Leech, G., Garside, R. & Bryant, M., 1994. CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*. Japan: Kyoto, pp. 622–628.

Leech, Geoffrey, 2005. Adding Linguistic Annotation. In M. Wynne, ed. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford, UK: Oxbow Books, pp. 17–29.

Leech, N.L. & Onwuegbuzie, A.J., 2007. A typology of mixed methods research designs. *Quality & Quantity*, 43(2), pp.265–275.

Lemieux, M. & Bordage, G, 1992. Propositional versus structural semantic analyses of medical diagnostic thinking. *Cognitive Science,* 16(2), pp.185-204

Leonard, J., 2004. Integrating Mathematics, Social Studies, and Language Arts with “A Tale of Two Cities”. , 35(3), pp.35–40.

Loftus, S.F., 2006a. *Language in clinical reasoning: using and learning the language of collective clinical decision making*. Ph.D THesis. Sydney: University of Sydney. Available at: http://hdl.handle.net/2123/1165 {Accessed 10 December 2011]. .

Loftus, S.F., 2006b. *Language in clinical reasoning: using and learning the language of collective clinical decision making*. University of Sydney.

Louis Lee, N. Y., Goodwin P.& Johnson-Laird G.2008 The psychological puzzle of Sudoku. *Thinking & Reasoning*:14 (4), 342–364.

Lubarsky, S. et al. (2011) Script concordance testing: a review of published validity evidence. *Medical education.* 45 (4), 329–338.

Lundgrén-Laine, H. & Salanterä, S., 2010. Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative health research*, 20(4), pp.565–75.

Lyle, J., 2003. Stimulated Recall: A Report on Its Use in Naturalistic Research. *British Educational Research Journal*, 29(6), pp.861–878.

Machado, J.L.M. et al., 2008. Self- and peer assessment may not be an accurate measure of PBL tutorial process. *BMC medical education*, 8, p.55.

Mamede, Silvia & Schmidt, Henk G, 2004. The structure of reflective practice in medicine. *Medical Education*, pp.1302–1308.

Mamede, Silvia, Schmidt, HG & Norman, GR, 2006. Innovations in problem-based learning: what can we learn from recent studies? *Advances in Health Sciences*, pp.403–422.

Mamede, Silvia, Schmidt, Henk G & Penaforte, J.C., 2008a. Effects of reflective practice on the accuracy of medical diagnoses. *Medical education*, 42(5), pp.468–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18412886 [Accessed September 2, 2010].

Mamede, Silvia, Schmidt, Henk G & Penaforte, J.C., 2008b. Effects of reflective practice on the accuracy of medical diagnoses. *Medical education*, 42(5), pp.468–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18412886 [Accessed September 2, 2010].

Mamede, Sílvia et al., 2010. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA : the journal of the American Medical Association*, 304(11), pp.1198–203.

Manning, G., 2009. Case 3.05 Benign or not benign ? *University of Nottingham:Graduate Entry Medical School*., pp.1–20. Nottingham:UoN.

Manning, G., 2008a. Case 3.01 Life is unpredictable. *University of Nottingham:Graduate Entry Medical School*; pp.1–19.Nottingham.UoN.

Manning, G., 2008b. Case 3.02 Be still my beating heart. *University of Nottingham:Graduate Entry Medical School*.Nottingham.UoN.

Manzar, S. & Al-Khusaiby, S.M., 2004. Crossword puzzle. A new paradigm for interactive teaching. *Saudi medical journal*, 25(11), pp.1746–7.

Marcum, J., 2012. An integrated model of clinical reasoning: dual-process theory of cognition and metacognition. *Journal of evaluation in clinical practice*, 18(5), pp.954–61.

Marshall, J.G. *et al*., 1993. A study of library use in problem-based and traditional medical curricula. , 81(3), pp.299–305.

Mattick, K. et al., 2008. Content specificity: is it the full story? Statistical modelling of a clinical skills examination. , 42(6), pp.589–599.

Maudsley, G, 1999. Do we all mean the same thing by “problem-based learning”? A review of the concepts and a formulation of the ground rules. *Academic medicine : journal of the Association of American Medical Colleges*, 74(2), pp.178–85.

Maudsley, G. 2011. Mixing it but not mixed-up: mixed methods research in medical education (a critical narrative review). *Medical teacher*, 33(2), pp.e92–104.

Mayring, P., 2004. Qualitative Content Analysis. *Forum: Qualitative Social Research*, 1(June).

McAleer, S. 2008 Clinical reasoning of Indonesian medical students as measured by diagnostic thinking inventory. *South East Asian Journal of Medical Education;*2 (1), 5.

McKimm, J., 2003. Curriculum design and development, e-learning for clinical teachers. London:Londond Deanary.

McKimm, J. & Barrow, M., 2009. Curriculum and course design. *British journal of hospital medicine (London, England : 2005)*, 70(12), pp.714–7.

Mclaughlin, D., 2008. Case 8.05 Twisted Sister.*University of Nottingham:Graduate Entry Medical School*. , pp.1–17. Nottingham:UoN

MCTES, 2006. *Tertiary Education in Portugal*, Reviews of National Policies for Education. OECD. Available from: t http://dx.doi.org/10.1787/104853273381 [Accessed 10 March 20111].

Medin, D.L., Altom, M.W. & Murphy, T.D., 1984. Given versus induced category representations: use of prototype and exemplar information in classification. , 10(3), pp.333–352.

Melina, 1997. *Computer-assited text analysis methodology in the social sciences*, Mannheim: *ZUMA-Arbeitsbericht 97/07*.

Melina & Zuell, 1999. A Review for Software for Text Analysis. *ZUMA-Nachrichten Spezial Band*.

Merriam, Sharan, B., 2009. *Qualitative Research: A Guide to Design and Implementation*, Chichester/Un: John Wiley & Sons, Ltd.

Merrill, S., 2006. Verbal Protocols: What Does It Mean For Research To Use Speaking As A Data Collection Tool? in M. C.-Deville, C. Chapelle, & P. Duff, Eds. *Inference And Generalizability In Applied Linguistics: Multiple Research Perspectives*. Amsterdam: John Benjamins.

MEC [no date] Acesso ao Ensiono Superior [Accessed 10 April 2011]. Available from: http http://www.sg.min-edu.pt/pt/acesso-ensino-superior/#

Migiro, S.O. & Magangi, B.A., 2011. Mixed methods : A review of literature and the future of the new research paradigm. , 5(May), pp.3757–3764.

Morgan, D.L., 2007. Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research*, 1(1), pp.48–76.

Mylopoulos, M. & Regehr, G., 2007. Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Medical education*, 41(12), pp.1159–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17986193.

Mylopoulos, M. & Regehr, G., 2011. Putting the expert together again. *Medical education*, 45(9), pp.920–6.

Nairne, J. 1992 Short-Term Forgetting of Order Under Conditions of Reduced Interference. *The Quarterly Journal of Experimental Psychology:* 52A (1), 241–251.

Neufeld, V. & Barrows, H., 1974. The“ McMaster Philosophy”: an approach to medical education. *Academic Medicine*. 49 (11);pp.1040-1050.

Norman, Geoff, 2002b. The long case versus objective structured clinical examinations. *BMJ (Clinical research ed.)*, 324(7340), pp.748–9.

Norman, G, 2004. Editorial – Theory Testing Research versus Theory-Based Research. *Advances in Health Sciences Education*, pp.175–178.

Norman, G.R., 2005. Research in clinical reasoning: past history and current trends. *Medical education*, 39(4), pp.418–27.

Norman, G R & Eva, K W, 2005a. Clinical experience and quality of health care. , 143(1), pp.85–6; author reply 86–7; discussion 87.

Norman, G R & Eva, K W, 2005b. Does clinical experience make up for failure to keep up to date? *Evidence Based Medicine*, 10(3), pp.A8–9.

Norman GR, 2006. Building on experience--the development of clinical reasoning. *The New England journal of medicine*, 355(21), pp.2251–2.

Norman, G R. et al., 2006. How specific is case specificity. *Medical Education*, 40, pp.618–623.

Norman GR,Young, M. & Brooks, L., 2007. Non-analytical models of clinical reasoning: the role of experience. *Medical education*, 41(12), pp.1140–5.

Norman, G R, Young, M. & Brooks, Lee R, 2007. Non-analytical models of clinical reasoning: the role of experience. *Medical education*, 41(12), pp.1140–5.

Norman GR et al., 2009. Iterative diagnosis. *BMJ*, 339(sep22 1), pp.b3490–b3490.

NLE [no date] University of Nottingham :Network Learning Environment [online] Available at:http://www.nle.nottingham.ac.uk/year1/index.php?session=2011/12.[Accessed 20 March 2012].

Lambert, C. et al. 2009. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol.* 4 (1), 7.

Oaksford, M., 2007. *Bayesian rationality : the probabilistic approach to human reasoning* , Oxford: Oxford University Press.

Offredy, M., Kendall, S. & Goodman, C., 2008. The use of cognitive continuum theory and patient scenarios to explore nurse prescribers' pharmacological knowledge and decision-making. *International journal of nursing studies*, 45(6), pp.855–68.

Ongoma, C., 2008. Case 2.02 To cap it all. *University of Nottingham: Graduate Entry Medical School*.Nottingham:UoN., pp.1–18.

Osgood, Suci & Tannenbaum, 1957. *The measurement of meaning.*, Urbana, US: University of Illinois Press.

OSLER, S.W. & F.R.S., 1909. *ALABAMA STUDENT: And Other Biographical Essays* 2nd ed., New York, USA: OXFORD UNIVERSITY PRESS AMERICAN BRANCH.

Osler, W., 1899. After Twenty-Five Years. , 28, p.823.

Osman, M., 2004. An evaluation of dual-process theories of reasoning. *Psychonomic bulletin & review*, 11(6), pp.988–1010.

Osterman, K., 1990. Reflective Practice: A new agenda for education. *Education and Urban Society*, 22(2), pp.133–152.

Page, G & Bordage, G, 1995. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. , 70(2), pp.104–110.

Neufeld, V. R. et al., 1981. Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical Education*, 15(5), pp.315–322.

Neuman, W.L., 2003. *Social Research Methods: Qualitative and Quantitative Approaches* J. Lasser, ed., Allyn and Bacon.

Neville, Alan J & Norman, Geoff R, 2007. PBL in the undergraduate MD program at McMaster University: three iterations in three decades. *Academic medicine : journal of the Association of American Medical Colleges*, 82(4), pp.370–4.

Neville, Alan J, 2009. Problem-based learning and medical education forty years on. A review of its effects on knowledge and clinical performance. *Medical principles and practice : international journal of the Kuwait University, Health Science Centre*, 18(1), pp.1–9.

Newman, 2004. Problem-based learning. In *Imaginative Curriculum Guide*. Higher Education Academy.

Newman, M., 2003. Higher Education Academy Imaginative Curriculum Guide Problem Based Learning. , (1989), pp.1–13.

Newman, M.J., 2005. Problem Based Learning: an introduction and overview of the key features of the approach. , 32(1), pp.12–20..

Niglas, K., 2004. *The combined use of qualitative and quantitative methods in educational research*. Tallinn Pedagogical University.

Norcini, J.J., 2003. Setting standards on educational tests. *Medical education*, 37(5), pp.464–9.

Norman, G R et al., 1985. Knowledge and clinical problem-solving. , 19(5), pp.344–356.

Norman, G R, 1989. Problem solving skills versus problem based learning. , 79(4), pp.307–310.

Norman, G R, *et al*., 1989. The development of expertise in dermatology.*J Med Ed*. 125(8), pp.1063–1068.

Norman, G R, Brooks, L R & Allen, S.W., 1989. Recall by expert medical practitioners and novices as a record of processing attention. , 15(6), pp.1166–1174.

Norman, G R & Schmidt, H G, 1992. The psychological basis of problem-based learning: a review of the evidence. , 67(9), pp.557–565.

Norman GR et al., 1999. The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning. *Cognition and Instruction*, 17(4), pp.433–448.

Norman, Geoffrey R. & Brooks, Lee R, 1997. The Non-Analytical Basis of Clinical Reasoning. *Advances in health sciences education : theory and practice*, 2(2), pp.173–184.

Norman, G R, 2000. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. , 75(10 Suppl), pp.S127–35.

Norman, G R & Schmidt, H G, 2000a. Effectiveness of problem-based learning curricula: theory, practice and paper darts. , 34(9), pp.721–728.

Norman, G R & Schmidt, H G, 2000b. Effectiveness of problem-based learning curricula: theory, practice and paper darts. *Medical education*, 34(9), pp.721–8.

Norman, Geoff, 2002a. Research in medical education: three decades of progress. *BMJ*, 324(7353), pp.1560–1562.

Page, Gordon et al., 1995. Developing key feature Problems and Examinations to Assess Clinical Decision-making Skills. *Academic Medicine*, 70(3), pp.194–201.

Pakman, M., 2000. Thematic Foreword : Reflective Practices : The Legacy Of Donald Schön. *Cyber net ics & Human Knowing*, 7(2), pp.5–8.

Papa, F.J., Shores, J.H. & Meyer, S., 1990. Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. , 65(9 Suppl), pp.S21–2.

Patel, V.L., Groen, G.J.J. & Frederiksen, C.H., 1986. Differences Between Medical Students And Doctors In Memory For Clinical Cases. *Medical Education*, 20(1), Pp.3–9..

Patel, V. L., Evans, D.A. & KAUFMAN, D.R., 1990. Reasoning strategies and the use of biomedical knowledge by medical students. *Medical Education*, 24(2), pp.129–136.

Patel, V L, Groen, G J & Arocha, J F, 1990. Medical expertise as a function of task difficulty. , 18(4), pp.394–406.

Patel, V L, Groen, G J & Norman, G R, 1991. Effects of conventional and problem-based medical curricula on problem solving. , 66(7), pp.380–389.

Patel, Vimla L, Groen, Guy J & Norman, Geoffrey R, 1993. Reasoning and Instruction in Medical Curricula. , 10(4), pp.335–378. Available at: http://www.informaworld.com/10.1207/s1532690xci1004_2.

Patel, Vimla L et al., 1994. Diagnostic Reasoning and Medical Expertise. In Academic Press, pp. 187–252.

Patel, V L & Arocha, J F, 1995. Cognitive models of clinical reasoning and conceptual representation. , 34(1-2), pp.47–56.

Patel, V. & Kaufman, D.R., 2000. Medical Expertise , Cognitive Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences*. New York, USA, pp. 9515–9518.

Patel, V. *et al.*, 2005a. Thinking and Reasoning in Medicine. In K. Novack, (eds). *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press.

Patel, V L *et al.*, 2005b. Knowledge integration and reasoning as a function of instruction in a hybrid medical curriculum. , 69(11), pp.1186–1211.

Patel, V L, Yoskowitz, N.A. & Arocha, J F, 2008. Towards effective evaluation and reform in medical education: a cognitive and learning sciences perspective. *Adv Health Sci Educ Theory Pract*, 14(5), pp.791–812.

Piaget, J., 1985. *The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development*, Chicago: University of Chicago Press.

PIAO, S. et al., 2005. Comparing and combining a semantic tagger and a statistical tool for MWE extraction. *Computer Speech & Language*, 19(4), p.19.

Pilnick, A. et al., 2010. Questioning the answer: questioning style, choice and self-determination in interactions with young people with intellectual disabilities. *Sociology of health & illness*, 32(3), pp.415–36.

Prince, K.J. et al., 2000. A Qualitative Analysis of the Transition from Theory to Practice in Undergraduate Training in a PBL-Medical School. , 5(2), pp.105–116.

Prince, K.J. et al., 2005. General competencies of problem-based learning (PBL) and non-PBL graduates. , 39(4), pp.394–401.

Quirk, M.E., 2006. *Intuition and metacognition in medical education: Keys to developing expertise*, Springer Publishing Company.

Radomski, N. & Russell, J., 2010. Integrated case learning: teaching clinical reasoning. *Advances in health sciences education : theory and practice*, 15(2), pp.251–64.

Rayson, Paul, 2000. Wmatrix : a web-based corpus processing environment [online]. Available at: http://ucrel.lancs.ac.uk/wmatrix.html [Acessed 29 March 2012]

Rayson, Paul & Garside, Roger, 2000. *Comparing corpora using frequency profiling.*, Hong Kong.

Rayson, Paul, 2002. *Matrix : A statistical method and software tool for linguistic analysis through corpus comparison*. Lancaster University.

Rayson, Paul, Berridge, D. & Francis, B., 2004. Extending the Cochran rule for the comparison of word frequencies between corpora. In G. Purnelle, C. Fairon, & A. Dister, eds. *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004)*. Louvain-la-Neuve, Belgium, pp. 1–12.

Rayson, Paul, 2008. From key words to key semantic domains. *International journal of corpus linguistics*, 13, pp.519–549.

Raupach, T. *et al.* 2009 Significant increase in factual knowledge with web-assisted problem-based learning as part of an undergraduate cardio-respiratory curriculum. *Adv Health Sci Educ Theory Pract.* 15(3);pp.349-56.

Regehr, G., 2004. Trends in Medical Education Research. *Learning*, pp.939–947.

Rich, S.K., Keim, R.G. & Shuler, C.F., 2005. Problem-Based Learning Versus a Traditional Educational Methodology: A Comparison of Preclinical and Clinical Periodontics Performance. , 69(6), pp.649–662.

Richmond, J. & Sowerby, P., 2004. The effect of familiarization time, retention interval, and context change on adults' performance in the visual paied-comparasion task. *Dev Psychobiol*.44(2).pp.146-55.

Rikers, Rm, Schmidt, Hg & Boshuizen, Hp, 2000. Knowledge Encapsulation and the Intermediate Effect. *Contemporary educational psychology*, 25(2), pp.150–166.

Rikers, Remy, Schmidt, Henk & Boshuizen, Henny, 2002. On the Constraints of Encapsulated Knowledge: Clinical Case Representations by Medical Experts and Subexperts. *Cognition and Instruction*, 20(1), pp.27–45.

Rikers, R.M.J.P. & Verkoeijen, P.P.J.L., 2007. Clinical expertise research: a history lesson from those who wrote it. , 41(12), pp.1115–1116. Available at: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2923.2007.02920.x.

Rimoldi, H.J., 1961. The test of diagnostic skills. , 36, pp.73–79.

Robertson, S.L., 2005. Re-imagining and rescripting the future of education: global knowledge economy discourses and the challenge to education systems. *Comparative Education*, 41(2), pp.151–170.

Rowan-Robinson, M., 2008. Case 8.04 Only the lonely. *University of Nottingham: Graduate Entry Medical School*. Nottingham:UoN.

Rowbottom, D.P., 2011. Kuhn vs. Popper on criticism and dogmatism in science: a resolution at the group level. *Studies in History and Philosophy of Science Part A*, 42(1), pp.117–124.

Salkind, N.J., 2010. *Encyclopedia of Research Design*, Kansas: SAGE Publications, Inc.

Savery, J. & Duffy, T., 1996. *Problem based learning: An instructional model and its constructivist framework*, Bloomington.

Savery, J.R., 2006. Overview of Problem-based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-based Learning*, 1(1).

Savin-Baden, M., 2000. *Problem-based learning in Higher Education: Untold Stories.*, Society for Research in Higher Education and Open University Press.

Savin-Baden, M., 2003. *Facilitating Problem Based Learning: Illuminating perspectives.*, Buckingham: Society for Research in Higher Education and Open University Press.

Savin-Baden, M., 2004. Understanding the impact of assessment on students in problem-based learning. *Innovations in Education and Teaching International*, 41(2), pp.221–233.

Schaufel, M. a, Nordrehaug, J.E. & Malterud, K., 2009. "So you think I'll survive?": a qualitative study about doctor-patient dialogues preceding high-risk cardiac surgery or intervention. *Heart (British Cardiac Society)*, 95(15), pp.1245–9.

Schmidt, H G, 1983. Problem-based learning: rationale and description. *Medical education*, 17(1), pp.11–16.

Schmidt, H G, Dauphinee, W.D. & Patel, V L, 1987. Comparing the effects of problem-based and conventional curricula in an international sample. *J.Med Ed*., 62(4), pp.305–315.

Schmidt, H G, Norman, G R & Boshuizen, H.P., 1990. Cognitive perspective on medical expertise: theory and implications. *Academic Medicine*, 65(10), pp.611–621.

Schmidt, H G *et al*., 1993. Influence of tutors' subject-matter expertise on student effort and achievement in problem-based learning. *Academic Medicine Journal of the Association of American Medical Colleges*, 68(10), pp.784–791.

Schmidt, Henk G. & Boshuizen, H.P., 1993a. On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), pp.205–221.

Schmidt, H G & Boshuizen, H.P., 1993b. On the origin of intermediate effects in clinical case recall. , 21(3), pp.338–351.

Schmidt, H G & Moust, J.H., 1995. What makes a tutor effective? A structural-equations modeling approach to learning in problem-based curricula. *AAMC Academic Medicine Journal of the Association of American Medical Colleges*, 70(8), pp.708–714.

Schmidt, H G et al., 1996. The development of Diagnostic Competece: Comparasion of a Problem-based, integrated and Conventional Medical Curriculum. *Academic Medicine*, 71(6), pp.658–664.

Schmidt, Henk G & Rikers, R.M.J.P., 2007. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical education*, 41(12), pp.1133–9.

Schmidt, H.G, Rotgans, J.I. & Yew, E.H.J., 2011. The process of problem-based learning: what works and why. *Medical education*, 45(8), pp.792–806.

Schön, D.A., 1983. *The Reflective Practitioner: how professionals think in action*, London, England: Temple Smith.

Schuwirth, L., 2009. Is assessment of clinical reasoning still the Holy Grail? *Medical education*, 43(4), pp.298–300.

Schuwirth, L.W.T. & Van Der Vleuten, C P M, 2005. Assessment of medical competence in clinical education. *Nederlands Tijdschrift Voor Geneeskunde*, 149(49), pp.2752–2755.

Schwartz, R.W. *et al.,* 1998. Assessing residents' clinical performance: cumulative results of a four-year study with the Objective Structured Clinical Examination. *Surgery*, 124(2), pp.307–312.

Scott, M., 1997. PC analysis of key words--and key key words. *System*, 25(2), pp.233–245.

Seale, C. et al., 2007. Media constructions of sleep and sleep disorders: a study of UK national newspapers. *Social science & medicine (1982)*, 65(3), pp.418–30.

Seale, C., Ziebland, S. & Charteris-Black, J., 2006. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. *Social science & medicine (1982)*, 62(10), pp.2577–90.

Sedlmeier, P. & Gigerenzer, G., 2001. Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130(3), pp.380–400.

Sefton, Gordon & Field, 2008. Teaching clinical reasoning to medical students. In Higg et al., eds. *Clinical reasoning in the health professions.* Elsevier.

Seigel, Shake, 2008. Case 2.05 Out of Africa.*University of Nottingham: Graduate Entry Medical School.*Nottingham:UoN

Shanteau, J., 1992. Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, pp.252–266.

Sibbald, M., Panisko, D. & Cavalcanti, R.B., 2011. Role of clinical context in residents' physical examination diagnostic accuracy. *Medical education*, 45(4), pp.415–21.

Da Silva, A.L. & Dennick, Reg, 2010. Corpus analysis of problem-based learning transcripts: an exploratory study. *Medical education*, 44(3), pp.280–8.

Simmons, B., 2010. Clinical reasoning: concept analysis. *Journal of advanced nursing*, 66(5), pp.1151–8.

Simon, Herbert Alexander & Chase, W.G., 1973. Skill in chess. *American Scientist*, 61(4), pp.394–403.

Skelton, J R & Hobbs, F D R, 1999a. Early reports Concordancing : use of language-based research in medical communication. *Lancet*, 353, pp.108–111.

Skelton, J R & Hobbs, F.D., 1999b. Descriptive study of cooperative language in primary care consultations by male and female doctors. *BMJ (Clinical research ed.)*, 318(7183), pp.576–9.

Skelton, J R, Murray, J. & Hobbs, F.D., 1999. Imprecision in medical communication: study of a doctor talking to patients with serious illness. *Journal of the Royal Society of Medicine*, 92(12), pp.620–5.

Skelton, John R, Wearn, A.M. & Hobbs, F D Richard, 2002. "I" and "we": a concordancing analysis of how doctors and patients use first person pronouns in primary care consultations. *Family practice*, 19(5), pp.484–8.

Smith, 2000a. Machine learning of well-defined thesaurus concepts. In Tan & Yu, eds. *Proceedings of the International Workshop on Text and Web Mining (PRICAI 2000)*. Melbourne, Australia, pp. 72–79.

Smith, 2000b. Machine mapping of document collections: the leximancer system. In *Proceedings of the Fifth Australasian Document Computing Symposium, DS*. Sunshine Coast, Australia.

Smith, 2003. Automatic extraction of semantic networks from text using leximancer. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Demonstrations - NAACL '03*, 4, pp.23–24.

Smith & Humphreys, M.S., 2006. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior research methods*, 38(2), pp.262–79.

Smith, M.K., 2001. David A. Kolb on experiential learning. The encyclopedia of informal education[online]. Available at: Ahttp://infed.org/mobi/david-a-kolb-on-experiential-learning/ [Accessed 12 January 2010].

Smith, C. S. 2008. A developmental approach to evaluating competence in clinical reasoning. Journal of Veterinary *Medical Education.* 35 (3), 375–381.

Sowa, 2000. *Knowledge representation: Logical, philosophical, and computational foundations.*, Pacific Grove, CA: Brooks Cole.

Sobral, D. T. 1995. Diagnostic ability of medical students in relation to their learning characteristics and preclinical background. *Medical Education.* 29 (4),pp.278–282.

Splinter, T. et al., 2007. Clinical Competence: General Ability or Case-specific? *Advances in Health Sciences Education*, 12(3), pp.299–314.

Stanovich, K. E., & Stanovich, P. J. 2010. A framework for critical thinking, rational thinking, and intelligence. In D. Preiss & R. J. Sternberg (Eds.), Innovations in educational psychology: Perspectives on learning, teaching and human development (pp. 195-237). New York: Springer.

Stanovich, K.E., 2010b. *Rationality and the reflective mind*, Oxford University Press, USA.

Stefano, F., Borsci, S. & Stamerra, G., 2010. Web usability evaluation with screen reader users: implementation of the partial concurrent thinking aloud technique. *Cogn Process*: 11(3);pp.263-72.

Stieger, S. et al., 2011. Diagnostic grand rounds: A new teaching concept to train diagnostic reasoning. *Eur J Radi*ol Jun;78(3);pp.349-52.

Struyf, E. et al., 2005. A new methodology for teaching clinical reasoning skills: problem solving clinical seminars. , 27(4), pp.364–368.

Stubbs, 1996. *Text and corpus analysis: Computer-assisted studies of language and culture.*, Oxford, UK: Blackwell Publishing Ltd.

Sys, F., 2003. Mindful Practice in Action (II): Cultivating Habits of Mind. *Health (San Francisco)*, 21(1), pp.11–17.

Tashakkori, A. & Teddlie, 1998. *Mixed methodology. Combining qualitative and qunatitative approaches*, London, England: Sage publications.

Tashakkori, A. & Teddlie, C., 2003. *Handbook of mixed methods in social & behavioral research* A. Tashakkori & C. Teddlie, eds., Thousand Oaks, California: Sage Publications, Inc.

Tavakol, M. & Dennick, Reg, 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, pp.53–55.

Taylor, C. & Gibbs, G.R., 2010. Preparing Data. *Online QDA Web Site*.Available at: http://onlineqda.hud.ac.uk/ [Accessed April 7, 2011].

Teddlie, C. & Yu, F., 2007. Mixed Methods Sampling: A Typology With Examples In V. L.P. Clark & J. W. Creswell,(eds). *Journal of Mixed Methods Research*, 1(1), pp.77–100.

Thomas, J. & Wilson, A, 1996. Methodologies for studying a corpus of doctor-patient interaction. In J. Thomas & M. Short, eds. *Using Corpora for Language Research*. London, England: Longman, pp. 92–109.

Thomas, P. a & Goldberg, H., 2007. Tracking reflective practice-based learning by medical students during an ambulatory clerkship. *Journal of general internal medicine*, 22(11), pp.1583–6.

Trochim, W.M.K., 2006. *Research Methods Knowledge Base 2nd Ed.*Available at: http://www.socialresearchmethods.net/kb/.{Acessed April 3, 2009]

Truell, A.D., 2003. Use of Internet Tools for Survey Research. *Information Technology, Learning, and Performance Journa*l , 21(1), pp.31–37.

UCREL, 2000. BNC2 POS-tagging Manual:Automatic POS-Tagging of the Corpus. *English*, pp.1–6. Available at: ucrel.lancs.ac.uk/bnc2/bnc2autotag.htm. [Accessed April 17, 2009].

Uebersax, J., 2010. Kappa Coefficients. *John Uebersax Enterprises LLC (c) 2000-2010*. Available at: http://www.john-uebersax.com/stat/kappa.htm [Accessed January 3, 2012].

University of Nottingham, 2011a. Graduate Entrance Medicine: Course Handbook. *Medicine*. Nottingham: UoN.

University of Nottingham, 2011b. Medical School Course Handbook:A100, A101, A300, A105 A104. *Medicine*, Nottingham:UoN.

UoN 2010.University of Nottingham:BMBs Medicine Degree webpage[online]. Available at: http://www.nottingham.ac.uk/ugstudy/courses/medicine/bmbs-medicine.aspx. .[Accessed 20 March 2012].

UoN 2010. University of Nottingham: Graduate Entry Medical School webpage[online]. Available from:http://www.nottingham.ac.uk/gem/prospective/graduateentrymedicine/courseoverview/clinical-training.aspx .[Accessed 20 March 2012].

University of Washington, , 2010. *Reporting Results of Common Statistical Tests in APA Format*, Washington.

Visschers-pleijers,A.. *et al.*, 2006. Analysis of verbal interactions in tutorial groups: a process study. *Medical education*, 40(2), pp.129–137.

Van Bruggen, L. *et al.* 2012. Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspectives on medical education.* 1 (4), pp.162–171.

Van der Vleuten, C P & Newble, D.I., 1995. How can we test clinical reasoning? , 345(8956), pp.1032–1034.

Walker, M., 1989. Analysing qualitative data: ethnography and the evaluation of medical education. , 23(6), pp.498–503.

Walker, S.E., 2003. Active Learning Strategies to Promote Critical Thinking. *Journal of athletic training*, 38(3), pp.263–267.

Wallis, S., 2001. Knowledge Discovery in Grammatically Analysed Corpora. *Knowledge Acquisition*, pp.305–335.

Waterhouse, H., 2009a. Case 3.03 It can wait until Monday.*University of Nottingham: Graduate Entry Medical School,* pp.1–17.

Waterhouse, H., 2009b. Case 3.04 Old and crusty. *University of Nottingham: Graduate Entry Medical School*, pp.1–19.

Watmough, S., Ryland, I. & Taylor, D.C., 2006a. Preregistration house officer skill and competency assessment through questionnaires. , 67(9), pp.487–490.

Watmough, S., Taylor, D. & Garden, A., 2006b. Educational supervisors evaluate the preparedness of graduates from a reformed UK curriculum to work as pre-registration house officers (PRHOs): a qualitative study. , 40(10), pp.995–1001.

Weber, R., 1990. *Basic content analysis.*, Newbury Park, CA: SAGE.

Wiley, J. & Jee, B., 2011. Cognition: Overview and Recent Trends. *Learning and Cognition*, 5, pp.245–250.

Wimmers, P.F. & Fung, C.., 2008. The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Medical Education*, 42(6), pp.580–8.

Wood, D., 2003. ABC of learning and teaching in medicine: Problem based learning. *BMJ: British Medical Journal*, 326(February).

Woods, N.N., Brooks, L R & Norman, G R, 2007a. It all make sense: biomedical knowledge, causal connections and memory in the novice diagnostician. *Adv Health Sci Educ Theory Pract*., 12(4), pp.405–415.

Woods, N.N., Brooks, L R & Norman, G R, 2007b. The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv Health Sci Educ Theory Pract*., 12(4), pp.417–426.

Woodward, C., Boud, D. & Feletti, G, 1997. What can we learn from programme evaluation studies in medical education. In London: Kogan Page.

Woollard, R.F., 2006. Caring for a common future: medical schools' social accountability. *Medical education*, 40(4), pp.301–13.

Yew, E.H.J. & Schmidt, Henk G, 2007. Process study of verbal interactions in problem based learning. ICT: Providing choices for learners and learning. Proceedings ascilite Singapore.pp.1119-1113

Young, M., Brooks, Lee R & Norman, G. R, 2007. Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical education*, 41(12), pp.1146–51.

Young, M., Brooks, Lee R & Norman, G.R, 2011. clinical reasoning The influence of familiar non-diagnostic information on the diagnostic decisions of novices. *Medical Education*, 45(4), pp.407–414.

Zavaleta-hernández, S. et al. 2011. Validation of the Script Concordance Test as an instrument to assess clinical reasoning of residents in pediatric emergency medicine in Mexico. 68 (5), 369–373.

Zhang, W., 2011. *Mixed Methods Embedded Design In Medical Education, Mental Health And Health Services Research: A Methodological Analysis*. Ph.D Thesis: University Of Nebraska.

Appendixes:

Appendix 1: Development of a new dictionary of medical terms

New dictionary of Medical Terms

The new dictionary introduced was an adaptation of the existing XML Medical Lexicon developed by Rothwell, Wheeler. & Nhàn, in 2006 (Rothwell *et al.*, 2006), and it is part of the Linguistics String Project (LSP), that began in 1965 at New York University. This project aimed to "implement an English language parsing program as a first step in the computer processing of the scientific literature. The goal was to facilitate the retrieval of specific information from texts in answer to queries by investigators." (Nhan, 2007).

The XML Medical Lexicon was designed as a set of tags, including 67,642 English lexical entries, to capture "medical, administrative and psychological elements [words] of a patient encounter" (Rothwell *et al.*, 2006). This classification system goes beyond the type of analysis required here, as its aim is to be able to capture and describe the entire content of medical documents, such as electronic personal records, appointments, and registrations, among others, and therefore includes many administrative categories that are not considered of relevance to our study. This is the type of tagging classification that is the basis of the Clinical Document Architecture format (CDA), that is at the core of the existing informatics systems used by health organisations.

The XML medical lexicon is divided into eight main classes, subdivided into 58 subcategories (see table below),; each word is tagged according to its linguistic and clinical properties.

Table AP1.1: XML medical lexicon eight main classes (Rothwell *et al.*, 2006)

Class of words	Description	Examples of subcategories
PATIENT	Words referring to patient	Anatomical area, specimen type occupation, nationality, traits physiological function.
TEST/EXAM	Words referring to tests or exams	Verbs of observation, test of specimen
TREATMENT	Words referring to patient treatments	Medical devices hospitals, clinics, physicians, staff complementary treatments
TIME	Words that refer to time relating a medical condition, procedure, consultation etc.	Location in time, beginning, termination
FINDINGS	Related with clinical findings	Indication of change, disease indicator word
EVIDENTIAL	Related with medical evidence	Uncertainty of finding; event has not yet occurred
OTHER 1	Group of words frequently used in medical context and not covered by any of the above.	Patient record terms

OTHER 2	Group of words frequently used in medical context and not covered by any of the above.	Chemicals

Development of the new dictionary

In order to be able to use the XML medical lexicon an adaptation was necessary. Firstly, the final purpose of the medical lexicon is not only for use in textual analysis, therefore its codings encompasses categories, such as administrative procedures, that are not of relevance. Secondly, the software used, Wmatrix2, does not use XML rules, therefore it was necessary to "translate" the XML format into the semantic and syntactic rules used by Wmatrix2 (USAS context rules). By identifying the subcategories that were not relevant for this research a total of six subcategories were eliminated. The words or multi-word structures in the remaining subcategories were extracted and re-scanned in order to identify non relevant words. The final list of words was re-coded using USAS lexicon rules (AP1.2). In order to achieve this format, lists of words had to be upload to the software, so that POS tagging was matching the parameters expected of it, then lists of annotated words extracted. To these lists the researcher manually added the semantic tags, and finally these were uploaded into the software and merged with the existing tagging systems (USAS and CLAWS). The accuracy of new classification was tested using extracted pieces of texts from the PBL.

Table AP1. 2: Example of USAS lexicon rules*1

*-foot	JJ	N3.7
*-full	JJ	N3.4
*-hour	JJ	T1.3

The final "new dictionary" upload had 43,053 words, plus 1365 multi-word expressions, distributed across the subcategories defined by the medical lexicon and the percentage total number of words in each subcategory is show by Table below. When the new dictionary was merged with the existing Wmatrix2 USAS tag system, to avoid duplication of tags (and then subsequent duplication of frequencies) the software automatically made a choice in favour of a pre-existing tag.

1. JJ corresponds to the syntactic tag: general adjective. N3.7, N3.4 andT1.3 are the semantic tags: Measurement: Length & Height, Measurement: Volume; Time: Period.

Table AP1. 1: "New dictionary" subcategories

	Wmatrix tag	Meaning
PATIENT	V11	words referring to patient
	V12	anatomical area
	V13	occupation, nationality, traits
	V14	physiological function
	V15	anatomical location relation
	V16	anatomical measure
	V17	palpated body part
	V18	body part
	V19	specimen type
	V191	verb with patient subject
TEST/EXAM	V21	verbs of observation
	V22	clinical exam, action
	V23	examination procedure
	V24	test of specimen
	V25	test variable
TREATMENT	V31	medical devices
	V32	hospitals, clinics, physicians, staff
	V33	complementary treatments
	V34	general medical management
	V35	treatment by medication
	V36	mode of administration
	V37	surgical procedure
TIME	V41	support class for prefix
	V42	beginning
	V43	termination
	V44	location in time
	V45	duration
	V46	time preposition
	V47	repetition
FINDINGS	V50	related with findings but not in any below
	V51	amount or degree

	V52	indication of change
	V52-	decrease
	V52+	increase
	V521	no change
	V53	**diagnosis**
	V54	disease indicator word
	V55	non-problematical
	V56	organism
	V57	patient response
	V58	test/exam result word
	V59	descriptive information
	V591	name of foods, dietary prescriptions
EVIDENTIAL	V60/v61	related with evidence but not in any below
	V61-	negation of finding
	V62	uncertainty of finding; event has not yet occurred
CONNECTIVE	V71	classifier verb
	V72	connects 2 information units
	V73	connects test and result
	V74	lab or measurement verbs
OTHER	V81	life stage
	V87	geographical names
	V88	racial or ethnic background
	V89	patient record terms
	V891	angles or positions of imaging
	V892	general terms that get classes from its modifiers
	V90	Chemicals
	V91	Units

Appendix 2: Analysis of PBL Sessions

Table AP2. 1: Key data on PBL sessions analysed

Session No	Session ID	Date	Module	Case	Session	Facilitator	Number of words with facilitator in clean file	Facilitators words (2)	Total number of words in file analysed	Duration of the sessions	Hours of recording (approx. values to hours)	Recording time (hh:mm:ss)	Word density per session/hour
1	GEM PBL 01c2.2	14/11/08	Respiratory	To cap it all	2	1	6953	103	6850	02:00	1.8	01:49:51	3806
2	GEM PBL 01c2.2	18/11/08	Respiratory	To cap it all	3	1	7010	79	6931	01:30	1.9	01:54:00	3648
3	GEM PBL 01c2.3	18/11/08	Respiratory	It's affecting my football	1	1	8885	120	8765	01:00	1.2	01:12:59	7304
4	GEM PBL 01c2.3	21/11/08	Respiratory	It's affecting my football	2	1	10105	0	10105	02:00	1.4	01:24:33	7218
5	GEM PBL 01c2.3	25/11/08	Respiratory	It's affecting my football	3	1	13273	73	13200	01:30	1.5	01:29:56	8800
6	GEM PBL 01c2.4	25/11/08	Respiratory	Can we fix it?	1	1	9604	73	9531	01:00	1	01:01:08	9531
7	GEM PBL 01c2.4	28/11/08	Respiratory	Can we fix it?	2	1	16209	258	15951	02:00	2.1	02:05:06	7596
8	GEM PBL 01c2.4	02/12/08	Respiratory	Can we fix it?	3	1	10487	394	10093	01:30	1.5	01:32:00	6729
9	GEM PBL 01c2.5	02/12/08	Respiratory	Out of Africa	1	1	8494	47	8447	01:00	1.2	01:12:31	7039
10	GEM PBL 01c2.5	05/12/08	Respiratory	Out of Africa	2	1	10532	40	10492	02:00	1.6	01:36:12	6558
11	GEM PBL 01c2.5	09/12/08	Respiratory	Out of Africa	3	1	5246	862	4384	01:30	1.4	01:23:50	3131
12	GEM PBL 01c3.1	09/12/08	Cardiovascular	Life is unpredictable	1	2	4734	450	4284	01:00	1	01:00:04	4284
13	GEM PBL 01c3.1	12/12/08	Cardiovascular	Life is unpredictable	2	2	16990	3073	13917	02:00	2	02:04:29	6959
14	GEM PBL 01c3.1	16/12/08	Cardiovascular	Life is unpredictable	3	2	12989	2908	10081	01:30	1.4	01:24:38	7201
15	GEM PBL 01c3.2	16/12/08	Cardiovascular	Be still my beating heart	1	2	9806	590	9216	01:00	1.1	01:06:00	8378
16	GEM PBL 01c3.2	19/12/08	Cardiovascular	Be still my beating heart	2	2	13766	1566	12200	02:00	1.8	01:48:00	6778
17	GEM PBL 01c3.2	13/01/09	Cardiovascular	Be still my beating heart	3	2	7270	2126	5144	01:30	1.3	01:18:00	3957
18	GEM PBL 01c3.3	13/01/09	Cardiovascular	It can wait until Monday	1	2	9859	80	9779	01:00	1.1	01:05:47	8890
19	GEM PBL 01c3.3	16/01/09	Cardiovascular	It can wait until Monday	2	2	15882	500	15382	02:00	2.1	02:06:00	7325
20	GEM PBL 01c3.3	20/01/09	Cardiovascular	It can wait until Monday	3	2	7414	936	6478	01:30	1.6	01:36:00	4049
21	GEM PBL 01c3.4	20/01/09	Cardiovascular	Old and crusty	1	2	5333	797	4536	01:00	1.6	01:34:06	2835
22	GEM PBL 01c3.4	23/01/09	Cardiovascular	Old and crusty	2	2	16586	8415	8171	02:00	2	01:05:00	4086
23	GEM PBL 01c3.4	27/01/09	Cardiovascular	Old and crusty	3	2	15434	3850	11584	01:30	1.6	01:36:00	7240
24	GEM PBL 01c3.5	27/01/09	Cardiovascular	Benign or not benign?	1	2	9941	2660	7281	01:00	1.3	01:18:00	5601
25	GEM PBL 01c3.5	30/01/09	Cardiovascular	Benign or not benign?	2	2	13982	2963	11019	02:00	1.6	01:35:45	6887
26	GEM PBL 01c3.5	02/12/08	Cardiovascular	Benign or not benign?	3	2	6998	1700	5298	01:30	1	01:00:50	5298
27	GEM PBL 02c8.4	02/12/08	Neurosciences	Only the Lonely	1	3	5676	133	5543	01:00	1.6	01:36:00	3464
28	GEM PBL 02c8.4	05/12/08	Neurosciences	Only the Lonely	2	3	5031	675	4356	02:00	1.9	01:53:46	2293
29	GEM PBL 02c8.4	09/12/08	Neurosciences	Only the Lonely	3	3	12186	1333	10853	01:30	1	01:00:30	10853
30	GEM PBL 02c8.5	09/12/08	Neurosciences	Twisted Sister	1	3	10400	1140	9260	01:00	0.9	00:54:00	10289
31	GEM PBL 02c8.5	12/12/08	Neurosciences	Twisted Sister	2	3	11463	1090	10373	02:00	1.5	01:30:00	6915

32	GEM PBL 02c8.6	16/12/08	Neurosciences	A Shaky Serve	3	3	2187	473	1714	01:30	2.3	02:08:58	745
33	GEM PBL 02c9.1	16/12/08	Integrative	Hard labour	1	4	6013	2467	3546	01:00	1	01:01:05	3546
34	GEM PBL 02c9.1	19/12/08	Integrative	Hard labour	2	4	6516	1346	5170	02:00	1.2	01:12:00	4308
35	GEM PBL 02c9.1	13/01/09	Integrative	Hard labour	3	4	8772	1888	6884	01:30	0.9	00:53:59	7649
36	GEM PBL 02c9.2	16/01/09	Integrative	Blood Brothers	2	4	9485	2434	7051	02:00	1.7	01:42:02	4148
37	GEM PBL 02c9.3	20/01/09	Integrative	More than just a rash	1	4	8402	1989	6413	01:00	1.3	01:18:00	4933
38	GEM PBL 02c9.3	23/01/09	Integrative	More than just a rash	2	4	6545	1902	4643	02:00	1.9	01:54:00	2444
39	GEM PBL 02c9.3	27/01/09	Integrative	More than just a rash	3	4	6589	3496	3093	01:30	1.5	01:30:00	2062
40	GEM PBL 02c9.4	23/01/09	Integrative	Put your foot down	2	4	10769	2850	7919	02:00	2.3	02:09:05	3443
41	GEM PBL 02c9.4	23/01/09	Integrative	Put your foot down	3	4	4823	1455	3368	01:30	0.7	00:42:00	4811

Table AP2 1: Sample characterisation data: gender, Entry year, Previous degrees, degree class, interview grade and GAMSAT score

Year	ID	Gender	Entry year	Age entry	Degree subject	Degree class	Interview grade	GAMSAT score
Year 1	ID0801	M	2008	24	Media technology	2.2	B	62
	ID0802	M	2008	25	Physics	2.2	C	62
	ID0803	M	2008	22	Biomedicine	2.2	C	60
	ID0804	M	2008	45	Zoology	2.1	B	66
	ID0805	F	2008	24	Cell biology	2.2	B	66
	ID0806	F	2008	29	Geography	2.2	B	62
	ID0807	F	2008	21	Biochemistry & Genetics	2.2	B	62
Year 2	ID0701	M	2007	24	Biomedical sciences	NA	B	65
	ID0702	M	2007	23	Music	1st	C	75
	ID0703	M	2007	25	Biology	2.2	A	67
	ID0704	F	2007	34	Biomedical sciences	2.1	A	62
	ID0705	F	2007	43	Emergency practitioner	2.1	B	61
	ID0706	F	2007	23	Natural sciences	2.2	B	62
	ID0707	M	2007	26	Maths	2.2	B	64

Table AP2 2: Code book and extracts of electronic content analysis 'themes'

	year 1			year 2		
Theme	**Definition**	**Connectivity ***	**Theme**	**Theme**	**Definition**	**Connectivity ***
1st sessions	Disease	This theme encompasses disease processes, chanism, and information provided on patient previous eases.	100%	Channels	technical discussion/biomedical principles/basic science	100%
	History	Theme encompasses discussions centred on patient dical history information.	83%	GABA	Technical discussion/biomedical principles/basic science	86%
	Causes	Theme encompasses causal relationships between the ɔrmation in the case presented and potential diagnosis, d pathophysiological mechanisms and processes of ease.	14%	Acid	Technical discussion/biomedical principles/basic science	23%
	Heart	Theme relates with cardiovascular system anatomy and ıction, with a high frequency of the word heart.	72%	things	Colloquialism related theme. Reflects the frequent use of word thing and its variations to replace technical minology or to replace explanations of processes.	23%
	GP	Theme related General practitioners (GPs). This theme compasses parts of the discussion where the students us on actions taken by the GP, information to be uested/provided by these professionals to the case sentation and actions to be taken by those such as errals, follow ups, treatments	14%	Because	This theme relates with words used to expressed ısation, e.g. words such as because/cause	8%
	Doing	Activity related theme, encompassing actions from ferent agents: The patient, actions from the practitioners olved in the case and actions of the group related with ir as discussion.	7%	Doing	Activity related theme, encompassing actions from ferent agents: The patient, actions from the practitioners olved in the case and actions of the group related with their discussion.	8%
	Night	Time related theme with high frequency of the ɔressions related with night.	6%	time	time related theme, often words such as e.g. hours are d	6%
				hydrogen	Technical discussion/biomedical principles/basic science	5%
				pH	technical discussion/biomedical principles/basic science	5%
				Serotonin	technical discussion/biomedical principles/basic science	5%
				potassium	technical discussion/biomedical principles/basic science	4%

* "connectivity score to indicate the relative importance of the themes (the most important is the top Theme at 100%). This score is calculated using the connectedness of concepts within that there to the other concepts in the dataset. This measures the importance of a theme within the dataset. Information retrieved from:https://hypermancer.leximancer.com/faq/display_and_output.html

	year 1			year 2		
	Theme	**Definition**	**Connectivity ***	**Theme**	**Definition**	**Connectivity ***
2nd Session	Heart	Theme relates with cardiovascular system anatomy and function, with a high frequency of the word heart.	100%	Stuff	Colloquialism related theme. Reflects the frequent use of the word thing and its variations to replace technical terminology or to replace explanations of processes.	100%
	Take	Action related 'theme' with high frequency of use of the word take in the context of the patient case	87%	Cells	technical discussion/biomedical principles/basic science	87%
	Time	time related theme, often words such as e.g. hours are used	86%	Motor	Theme relates with movement (muscular anatomy and function) with a high frequency of the word heart.	74%
	Pressure	technical discussion/biomedical principles/basic science	12%	Things	Colloquialism related theme. Reflects the frequent use of the word thing and its variations to replace technical terminology or to replace explanations of processes.	67%
	Remember	High use of expressions that indicate need o remember/revise facts and information	10%	Ones	Colloquialism related theme. Reflects the frequent use of the word thing and its variations to replace technical terminology or to replace explanations of processes.	34%
	Goes	Action related theme, related with the patient's actions.	9%	DNA	technical discussion/biomedical principles/basic science	32%
	Doctor	High use of the word doctor, associated with doctor's responsibilities	5%	Makes	Theme related with impact, consequences of actions e.g:"it makes the cells react causing them to"	30%
				Remember	High use of expressions that indicate need o remember/revise facts and information	25%
				Called	Theme associated with naming , in most utterances related with common names(e.g. "what are these cells called?")	7%
				Thought	Theme expresses suggestions, opinions, expression of thoughts often introduced to the group as the following except " other thing I thought yesterday"	6%
				Virus	Technical discussion/biomedical principles/basic science	5%
				sure	Groups the words the demonstrate agreement. Also indicates colloquialism. e.g. sure being the most frequent in expressions such as: "yes, sure!"	4%

* "connectivity score to indicate the relative importance of the themes (the most important is the top Theme at 100%). This score is calculated using the connectedness of concepts within that there to the other concepts in the dataset. This measures the importance of a theme within the dataset. Information retrieved from:https://hypermancer.leximancer.com/faq/display_and_output.html

	year 1			Year 2		
	Theme	**Definition**	**Connectivity ***	**Theme**	**Definition**	Connectivity *
3rd sessions	things	Colloquialism related theme. Reflects the frequent use of the word thing and its variations to replace technical terminology or to replace explanations of processes.	100%	Different	Theme associated with search or identification of differences e.g. "What about a different system? or the upper respiratory tract?"	100%
	lung	Theme encompassing related to lung function and the respiratory system	55%	Nerves	Technical discussion/biomedical principles/basic science	87%
	heart	Theme relates with cardiovascular system anatomy and function, with a high frequency of the word heart.	46%	Blood	Technical discussion/biomedical principles/basic science	58%
	day	Time related theme with high frequency of the expressions related with day (e.g. "during the day")	27%	Positive	Technical discussion/biomedical principles/basic science	44%
	uncertainty	Theme associated with expression of doubt, uncertainty or used to ask fro clarification.	5%	Atypical	Different and atypical situations, mainly in relation to atypical patient presentations	26%
	sure	Groups the words the demonstrate agreement. Also indicates colloquialism. e.g. sure being the most frequent in expressions such as: "yes, sure!"	4%	People	Theme associated with generic trends in groups or populations e.g."Yeah that's the recommended treatment, but you should only have that in a stroke centre with people[health professionals] that know what they're doing"; "usually in about 70% of people"	18%
				Food	generic discussion about food/some case related material	13%
				Infection	technical discussion/biomedical principles/basic science	12%
				Uncertainty	Theme associated with expression of doubt, uncertainty or used to ask fro clarification.	8%
				Thought	Theme expresses suggestions, opinions, expression of thoughts often introduced to the group as the following example " other thing I thought yesterday"	8%
				Doing	Activity related theme, encompassing actions from different agents: The patient, actions from the practitioners involved in the case and actions of the group related with their as discussion.	6%
				Question	Theme related with the use of words such as expression of inquiry or interrogative sentences.	1%

* "connectivity score to indicate the relative importance of the themes (the most important is the top Theme at 100%). This score is calculated using the connectedness of concepts within that there to the other concepts in the dataset. This measures the importance of a theme within the dataset. Information retrieved from:https://hypermancer.leximancer.com/faq/display_and_output.html

Appendix 3: Questions/Marks/Models Answers Table

	Question	Marks	Type of question	Type of marking	Model answer Case 1	Model answer Case 2
1	Summarise the patient information you have just heard (limited space 300 words).	10 marks (1 mark per key feature identified from list)	Short-answer	Manual	Key features: 60 years old Male Chest pain Pain duration 1hour Pain intensity:intense Pain radiation: back/stomach Sudden onset/Ambulance 999 Associated sob Pales, sweaty Confused/anxious	10 key features (1 each) 17 year old female Dehydration Difficulty concentrating, tiered Weak Vomiting Diarrea 2 days stomach hurts SOB
2	List here the main clinical problems:	3 marks per any 3 problems (from the list)	Short-answer	Manual	List: Chest pain Back pain Stomach pain Pain all over SOB Confusion	List: SOB Stomach pain Vomiting Diarrea Dehydration2
3	If you could choose the course of action to follow, what would you like to do immediately?	5 marks correct; 4marks alternative	Multiple choice question	Automatic (software)	Correct option take history; Alternative is Order treatment (only to be considered if treatment is O2)	Correct option take history; Alternative is Order treatment (only to be considered if treatment is fluids)
4	Please explain why would you take that course of action.	5 marks for good explanation;3 marks for vage explanations (e.g. need more information)	Short-answer	Manual	The correct course of action at this point would be to take history, as the information provided so far is not yet enough to do a diagnosis of this patient. The option start treatment was also considered correct, as long as the treatment is high flow O2. Due to the small amont of information you have at this stage, any other treatment would be represent an unnecessary risk for the patient.	Take history, EXPLANATION: not have enough information, patient is able to speak therefore is important to get more information in order to make a decision on the course of action. Start treatment EXPLANATION: Give her fluids to deal with the dehydration while making a diagnosis
5	Essential questions	1 mark per correct question selected	Multiple choice question	Automatic (software)	3$7$8	not included in this case
6	Importantbutnotverydiscriminatory	1 mark per correct question selected	Multiple choice question	Automatic (software)	1$5$6	not included in this case
7	Least relevant questions	1 mark per correct question selected	Multiple choice question	Automatic (software)	2$4$9	not included in this case

8	If you would like to ask any other questions to the patient, please write them here:	5 marks any 2 relevant questions (containing elemnets of this case); 2.5 vague "SOCRATES" was given 1.5 as it is relevant but it is a general mnemonic for a series of questions not a especific question.	Short-answer	Manual	Clarification about the pain (mainly Radiation, Quality, Quantity, Duration, Frequency, Aggravating Factors,Relieving Factors or Associated Symptoms). Clarification on medical history,familiar history. Questions should either related with a presented feature (clarification) or be clearly directed to a feature that might be of relevance for this particular patient. Questions that are less discriminatory are of a lower priority at this stage; remember the context here is the A&E and your patient is in pain, doesn't yet have a diagnosis, and may need fast treatment. 5 marks (any 2 relevant questions, 2.5 vague questions in any 2 relevant) "SOCRATES" was given 1/2 as it is relevant but it is a general mnemonic for a series of questions not a especific question.	Clarification about the pain, mainly Radiation, Quality, Quantity, Duration, Frequency, Aggravating Factors,Relieving Factors or Associated Symptoms. Clarification on familiar history, health history, can be considered others: Family history Bowel movements Important aspect is to relate the questions to the clinical presentation.
9	Diferential Diagnsosis Matrix	3 marks for just correct (although is the correct answer you should have, at this stage more option in mind) 4 marks for 2 plausible answers (not most appropriated column/row) 5 marks for 1 plausible answers (not most appropriated column/row) + MI (high-high) 6 marks 1 right answers and a plausible answer (not most appropriated column/row) 7 for 2 right answers, without the corect 8 marks for 2 right answers, one being the correct diagnosis	Matrix	Manual	List of differentials considered: Myocardial infarction, ischemic heart disease, coronary heart disease acute coronary syndrom, Aortic dissection Ischaemic heart disease Pulmonary embolism Dissecting aortic aneurysm abdominal aortic aneurysm (AAA) Pulmonary embolism Cholecystitis Pericarditis Angina Others maybe considered correct in particular cases	LIst of differentails considered: DKA, GI pathologies, PANCREATITIS, poor diabetic control, infection, hypoglycæmia, Inflammatory bowel disease, appendicitis, Eating disorder food poising,alcohol poisoning, salmonellosis, campylobacter, tropical disease, alcohol poisoning, gastro-oesophageal reflux, coeliacs disease (GORD), stress, Endocrine: thyroid disease
10	Marks Select from the list 3 results you would expect to find in the physical examination?	1 mark per correct selected answer	Multiple choice question	Automatic (software)	Tachycardia/Tachypnoea;Absent bowel sounds;Upper abdominal rebound tenderness	Dry mucous membranes, Strong smelling breath,Tachycardia/Tachypnoa

11	Diferential Diagnsosis Matrix 2	3 marks for just correct (although is the correct answer you should have, at this stage more option in mind) 4 marks for 2 plausible answers (not most appropriated column/row) 5 marks for 1 plausible answers (not most appropriated column/row) + MI (high-high) 6 marks 1 right answers and a plausible answer (not most appropriated column/row) 7 for 2 right answers, without the corect 8 marks for 2 right answers, one being the correct diagnosis	Matrix	Manual	List of differentials considered: Myocardial infarction, ischemic heart disease, coronary heart disease acute coronary syndrom, Aortic dissection Ischaemic heart disease Pulmonary embolism Dissecting aortic aneurysm abdominal aortic aneurysm (AAA) Pulmonary embolism Cholecystitis Pericarditis Angina Others maybe considered correct in particular cases	LIst of differentails considered: DKA, GI pathologies, PANCREATITIS, poor diabetic control, infection, hypoglycæmia, Inflammatory bowel disease, appendicitis, Eating disorder food poising,alcohol poisoning, salmonellosis, campylobacter, tropical disease, alcohol poisoning, gastro-oesophageal reflux, coeliacs disease (GORD), stress, Endocrine: thyroid disease
12	Chose clinical investigations	1 mark per correct option selected total 5 marks	Multiple choice question	Automatic (software)	A, B, D, F and N will be the correct options for this questions. SCT I (Chest-X Ray) 0.5 Although other options maybe given partial marks considered if consistent with your main diagnosis hypothesis (as indicated on question 13 and/or 17). Chest-X ray was For example: 0.5 for C if colicystitis 0.5 for I if perfurated ulcera 0.5 for F and K if AAA 0.5 for M if gastric ulcera	A, B, D, E and G will be the correct options for this questions. Although other options maybe given partial marks considered if consistent with your main diagnosis hypothesis (as indicated on question 11 and/or 15).
13	Write here 2 results you expect to find in the clinical investigations	2 marks (per any 2 correct answers)	Short-answer	Manual	Typical results to expect are the following: Cardiac markers elevated (Troponin and CK-MB) elevated ECG abnormal	Typical results to expect are the following: High Capilary glucose Abnormal urinalysis or Ketones in urine

14	Scipt concordance questions	1 mark by correct answer (same selection as the experts)	Script conrdance	Manual	...Raised troponin and CK-MB > more/extremly likely ...Chest-X ray normal> more/extremly likely ...Faecal blood test positive for blood > not effect/less likely ...High O2 and low PaCO2 >would not affect ...ECG shows a raised ST segment >more likely ...C-reactive protein (hs-CRP test) is higher than 2.4 mg/l (normal range: 0-1.0 mg/l) >more likely [0.8 and 0.5 can be given depending on the answer and how it relates with the main diagnosis (Q11 and/or Q15)]"	...Urinalysis reveals normal levels of glucose and ketones, Extremely unlikely/unlikely; ...Chest X-ray is normal (Not affect); ...Slight increase in Hb (17g/dl)(Not affect); ...Serum amylase is normal (more likely)
15	Based on the entire history, examination and investigations, what do you consider is now the most likely diagnosis?	2 marks	Short-answer	Manual	MI	DKA
16	Why do you think that is the most likely diagnosis?(Provide reasons for your answer to the previous question)	2 marks per at least 2valid reasons	Short-answer	Manual	Justification must include information given in at least 2 of the following parts of this case: Clinical investigations, history, physical examination, clinical presentation.	Justification must include information given in at least 2 of the following parts of this case: Clinical investigations, history, physical examination, clinical presentation.
17	Scipt concordance questions 2	1 mark by correct answer (same selectic	Script conrdance	Automatic (sc	New informtion will not affect the diagosis or Diagnosis will became more likely	not included in this case

Appendix 4: Focus Group Summary

Positive aspects of the CRT

Good structure and sequence of questions. The overall opinion of the three participants was very positive. The participants thought the CRT was very useful way to prepare them for clinical practice but also a useful tool to revise for exams. The structure and sequence of questions were identified as very pertinent and highly clinically relevant. Overall *"It is fun [to answer these cases]"* (participant 1).

"You have to train to listen to patients" (participant 2). The audio case presentation and the clear introductory context were also very positive aspects mentioned by all. Participants also mentioned, as particularly positive, the question where they were asked to summarise the patient information as a useful training experience for practice: *"[in clinical practice] you can't ask a patient again, and again and again to repeat"* (participant 3).

"That's [the cases] what you get out there [in the hospital]" (participant 3). The clinical content of the cases was rated as being realistic and related to what they are expected to see in daily practice. The participants agreed that the cases presented were what they are expected to be able to deal with in their clinical rotations and once they graduate.

Reasoning or factual knowledge? The participants felt it was difficult to distinguish between the knowledge and the reasoning process, and thought that both were being assessed simultaneously in the cases. This is understood as a positive comment, as the clinical reasoning process is highly dependent on knowledge and therefore it was expected that the students felt that both were being used to answer the cases.

Participants were asked if, in their opinion, having textbooks would help them answer the cases. ***"It won't help much. I think you would still have to think a lot to get to the answers. These answers do not come in text books"*** (participant 2). Although two of them agreed that a book would not be helpful, one of the participants actually said that they could be of use, although recognising that they would not provide answers to the questions.

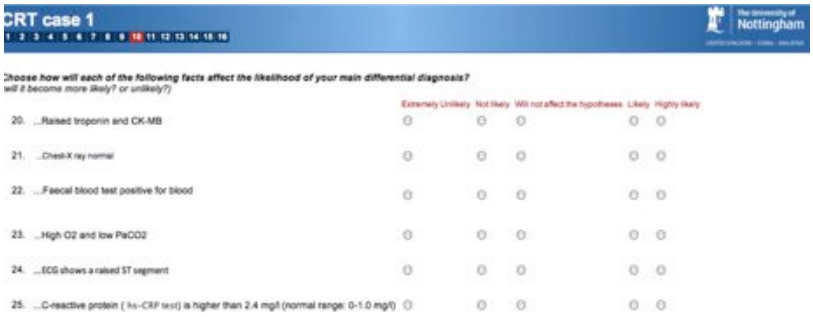
Mixed comments

More information. Initial information about the test format, its aim and how to answer the questions was provided in the participant's information sheet, an introductory screen and reinforced by the researcher. Although mentioning that the introduction was clear, the participants agreed that, after the first case it was easier to answer the second case, as they knew what to expect from the test. This was due to the format being, in their opinion, clearly different form the usual test format they are used to in their assessments. A reason for this may also be the fact that this is a sequential and unidirectional test; therefore the students could not see all the questions simultaneously, as they would in other assessments.

Regarding the level of difficulty there was some discussion among participants. All agree that these should in principle be moderate/easy cases once they had more clinical experience (final-year) and that the cases would be challenging if they had to answer them one year before (that is without exposure to clinical practice). However, while for one, the first case of myocardial infarction, (MI) was felt to be the easiest and the diabetic ketoacidosis (DKA) case to be the most challenging from the four cases, the other two found the MI case to be "tricky" due to the combination of chest and back pain. All agreed that the chronic obstructive pulmonary disease and the gastroenteritis by *campylobacter* case, should not be too difficult, especially for someone at the end of their degree.

Two question formats were identified as the most difficult or challenging, the script concordance type of question, and the question where participants were asked to predict the results of clinical investigations. Regarding the first, the main reason identified for its difficulty was the scale of response as it possible to see by Figure 6. 1 below. In fact it is possible that the students had not had any previous contact with SCT type of questions, therefore perceiving them as difficult. Regarding the second question, its difficulty was due to the novelty of this question, as according to them they are usually asked to interpret clinical investigations, not predict their results. However, this was considered a positive training aspect of the CRT cases.

Figure 6. 1: Script concordance type question



CRT case 1

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

The University of Nottingham

Choose how will each of the following facts affect the likelihood of your main differential diagnosis?
(will it become more likely? or unlikely?)

	Extremely Unlikely	Not likely	Will not affect the hypotheses	Likely	Highly likely
20. ...Raised troponin and CK-MB	O	O	O	O	O
21. ...Chest-X ray normal	O	O	O	O	O
22. ...Faecal blood test positive for blood	O	O	O	O	O
23. ...High O2 and low PaCO2	O	O	O	O	O
24. ...ECG shows a raised ST segment	O	O	O	O	O
25. ...C-reactive protein (hs-CRP test) is higher than 2.4 mg/l (normal range: 0-1.0 mg/l)	O	O	O	O	O

"What would you like to do next?" Participants 2 and 3 mentioned that they were not certain about the question "what would you like to do next?", as they would like to have had more information before deciding or more options that combined more than one option (e.g. take history and call cardiologist team). Therefore, although perceiving it as a good question, they felt limited by the options as, in their opinion, it was difficult to choose only one between them. However, participant one did not feel the same way about this question.

Improvements needed

Initial information could be clearer. Although the initial information provided was perceived as useful and well presented, participants mentioned that it was not clear enough in the introduction that the test was unidirectional, leading to them not paying enough attention when listening to the case. According to them this would not have happened if the instructions were more explicit.

More questions on patient history. Participants felt that there was not enough information available from the history of the patient and they all agreed that they would like to be able to ask the patient more questions apart from the ones presented in the test.

Give normal values. Participants felt they would benefit if they were given normal values for clinical investigations. Although one actually thought that this should be something they ought to know beforehand on clinical practice.

Include EGC's, X-Rays, and others for interpretation. One participant suggested that more media elements could be added to the cases, such as ECGs or X-rays.

Technical issues. Some technical issues were identified, mainly a word missing on some of the screens, some of the audios were too "robotic" (COPD and MI), and some fonts were too small to read properly in the introduction.

Time and suggestions for data collection. The students anticipated that to answer all the cases, if they were being tested, they would need more than one hour or 1 hour and 30 minutes. They also mentioned that they have very busy schedules and it would be difficult to

find suitable times for the students to be physically in the medical schools to take the CRT. This was especially true for final-year students, as they are in clinical placements and not very often in the medical school building.

Appendix 5: CRT Case Development Guide and CRT Cases 1&2

The University of Nottingham

CLINICAL REASONING TEST NEW CASE GUIDE

Ana L. Da Silva, Prof. Reg Dennick, Dr. Danny McLaughlin

- ✓ Case overview
- ✓ Case summary (summary of the case key features)
- ✓ Patient Initial Presentation
- ✓ Selection of information I
- ✓ Selection of information II
- ✓ Differential diagnosis I
- ✓ Differential diagnosis II
- ✓ Clinical investigations
- ✓ Script concordance Question
- ✓ Final Diagnosis
- ✓ Treatment

The present document outlines the information needed in order to create new cases for the Clinical Reasoning Test

Case overview

1. Who is this aimed at (target)

2. Educational context (curriculum phase/module)

3. Type of assessment (Progress test/Formative/summative)

4. Delivery (own time/timetabled)

Case summary (summary of the case key features)

Patient Initial Presentation (screen 1)

1. Demographics (Name, age, gender, nationality, profession...)

2. Level of responsibility of the student (eg. F1, GP....)

3. Location (GP surgery, A&E,...)

4. Description of the patient (describe any signs detectable when the clinician encounters the patient. Mention general appearance. Eg. looks weak, has the patient smells to alcohol...)

5. Description of the situation (describe how the patient come to meet the clinician. Important aspects to mention here are for example if the patient come by ambulance, was it an internal patient in a ward...)

6. Transcript for audio/video file (This text should represent the patient complains, including its signs, symptoms, concerns and other elements that make this a realistic piece of spoken text. Do not provide concise summaries of signs and symptoms, patients are often imprecise, vague and do not use medical jargon. Language type and level should be coherent with patient educational and socio cultural background.)

7. Other useful elements

Selection of information I (screen 2)

1. Key point to be identified by the students from the above presentation

    a. Number of marks per each correct identified
    b. Total number of marks given to the question

2. List of clinical problems to be identified by the students

    a) Number of marks per each correct identified
    b) Total number of marks given to the question

3. Correct course action at this moment (think student level) and its justification.
   Consider other possible answers (less correct) dependent on the explanation given.

History	Justification:
Physical Examination	Justification:
Clinical investigations	Justification:
Start treatment	Justification:

a. Total number of marks given to the questions

Selection of information II (screen 3)

1. Select 3 questions for each of the categories below according to their discriminatory value in this case and provide the answers the patient would give to the questions

	Questions	Patient Answers (screen 4)
Essential questions		
Important but not very discriminatory		
Least relevant questions		
Other distracter questions (if necessary)		

2. Other possible valid questions

    a. Total number of marks given to the questions

Differential diagnosis I (screen 3)

1. Provide the answers to the questions asked previously

2. Should the students be able to do a differential at his point?
   a. If yes use the table to write at least 3 possible answer for each category

		Likelihood of diagnosis	
		High	Low
Seriousness	High	1. 2. 3.	1. 2. 3.
	Low	1. 2. 3.	1. 2. 3.
Other options to be considered:			

   b. Total number of marks given to the question

Physical Examination (screen 4)

1. List 3 results you would expect to find in the physical examination
   a)
   b)
   c)

2. List 4 results you would not expect to find
   d)
   e)
   f)
   g)

   a. Some answers may be dependent on the answers to the DDx matrix (hand marking may be necessary)
   b. Total number of marks given to the question

3. List results from the physical examination (screen 5)

- BMI:20
- Heart rate:
- BP:
- Respiratory Rate:
- Chest examination:
- Abdominal examination:
- .
- .

Differential diagnosis II

- If in your case the Physical examination result are expected to have a impact on the DDx matrix write the new DDx, if not use the previous one.
- Marks could be stricter here, as there is more information available.

Seriousness		Likelihood of diagnosis	
		High	Low
	High	1. 2. 3.	1. 2. 3.
	Low	1. 2. 3.	1. 2. 3.
Other options to be considered:			

a. Total number of marks given to the question

Clinical investigations (screen 6)

	Name of clinical invest.	Tests included	Patient exposure to risk (1 low to 4 high)
A	Blood tests (Full Blood Count)	Haemoglobin (Hb); Red Blood (Cell) Count (RBC); Haematocrit (HCT); Mean Cell Volume (MCV); Mean cell haemoglobin (MCH); Mean cell haemoglobin concentration (MCHC); White Blood (Cell) Count (WBC); Platelets (PLT); Neutrophils (NEUTS); Lymphocytes (LYMPH); Monocytes (MONOS);Eosinophils (EOS);Basophils (BASOS);Erythrocyte sedimentation rate (ESR);Reticulocytes	1
B	Urea and Electrolytes	Creatinine (CREA) ; Glucose (GLU) [Plasma] Potassium (K) ; Sodium (NA) ;Urea (UREA)	1
C	Blood Sample (bioq.)	Albumin (ALBG); Alkaline Phosphatase (ALP) ; Alanine Transaminase (ALT); Aspartate Transaminase (AST) ; Bicarbonate (BIC) HCO3Bilirubin (BILI) ; Calcium (CAL); Calcium Corrected (CALA);Chloride (CL); Creatinine (CREA); Gamma-Glutamyl Transpeptidase (GGT); Glucose (GLU) [Plasma]; Osmolality (OSMS); Potassium (K); Protein, Total Serum (TP); Sodium (NA); Urea (UREA).	1
D	Arterial Blood Gas (ABG)	Bicarbonate (BIC) HCO3; pH; PCO2; PO2; Base Excess	2
E	Capillary glucose	.....	1
F	Troponin and CK-MB (cardiac markers)	Creatine Kinase (CK); Creatine Kinase (CK-MB) testLactate Dehydrogenase (LDH); Myoglobin (MYO) [Urine, Random]Troponin T (TnT) [Serum]	1
G	Urinalysis	Blood; Protein; Glucose Ketones; Nitrates; White cells	1
H	Stool Culture	----------	1
I	Chest X-ray	----------	2
J	Abdominal X-ray	----------	2
K	Abdominal CT scan	----------	3
L	Coronary Angiogram	----------	4
M	Upper abdominal Endoscopy	----------	4
N	ECG	----------	1

1. List the CI to ask (use the table) for this patient and its results

Clinical Investigation	Results (screen 9)	Can you expect a student to predict this result? (screen 6)
		yes/no
		yes/no
		yes/no
		yes/no

a) Number of marks per each correct identified
b) Total number of marks given to the question

Script concordance question (screen 8)

1. List a few facts that could have an impact on the diagnosis at this stage
2. List a few that are irrelevant
3. Ask experts to rate their impact on a 5 points likert scale

Facts	Experts 1	Experts 2	Experts n	Decision
Non relevant fact				
Relevant fact				

a. Total number of marks given to the question

Final Diagnosis (screen 9)

1. List all the CI test not preformed and all the CI results (use realistic images if possible)

NHS Trent Region

FULL BLOOD COUNT

Name	Value	Units	Normal Range

2. What is the most likely diagnosis?
    a. Total number of marks given to the question

3. Model answer for a good justification?
    a. Total number of marks given to the question

Other options

Add more uncertainty (Script concordance test)

- Test students' ability to review their previous decisions (eg)
- Test ability to apply their knowledge of drugs an medications (eg)

Treatment

- Consider how important it is to test this accordingly to the students' level (curriculum phase).
- Knowing what treatments are available is a matter of knowledge not reasoning.
- If, for example, there are interactions with other medications that may require changes on prescriptions, if certain medication s do not have the expected outcome or the patient experience side effects, then reasoning will be involved.

Research Questions

- Use of additional resources
- Confidence
- Knowledge about a disease
- CR strategies & readiness to decide/act

CRTCP3

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Clinical Reasoning Test CP3

Dear Student:

The case you will encounter in this progress test is part of a study of the development of clinical reasoning in undergraduate Medical Education. This study, carried out by Ana Da Silva, under the supervision of Professor Reg Dennick aims to explore the impact of different educational strategies on the development of clinical reasoning in undergraduate medical education.
If you had already taken part on this study in its pilot phase you

The test: What to expect?
Here you will find a clinical case, starting with patients presentation, followed by several sequential new informations about the case and questions.

This is an **unidirectional test**, therefore once you select to go to the next screen you **CANNOT go back** to the previous one, or have access to the information displayed there.

The test will be **active for a week**, as the previous one's. However the **test must be completed all at once**. Meaning once you started it you must complete it, otherwise your answer will be not be recorded and we won't be able to provide you with your feedback.

The questions: What to expect?
As this test was developed to test your clinical reasoning while allowing a deeper understanding of the clinical reasoning strategies use, some questions will not be counted in your final mark. These questions are identified by a *"note: research question"* and will be used only for research purposes and to give you a more complete feedback on your results.

Feedback: What to expect?
Due to the nature of this progress test the feedback will have 3 main components:

a) Model answers
b) Your mark. Your mark will give you an idea of how close did your answers related with the model answers.
c) Your Clinical Reasoning "profile". This "profile" will show which parts of the clinical reasoning process you are stronger and in which you can improve in these two cases.
This profile will only be available if you agree to take part in the study allowing us to further process your data, and all the information will be **treated as CONFIDENTIAL** and will **NOT BE SHARED** be shared with any third parties.

What we expect form you?
To be interested and use this as an opportunity to further develop your skills.Answer the test, as good as you can, so the feedback we give is as authentic as possible.

If you decide to use books or **other sources** of information, just **let us know** (final questions). This will **not** have an impact **on your mark** but will be extremely important for our research.
If you decide to challenge yourself and do the test without external sources of information... **GREAT!!** Your feedback will clearly be much more authentic.

THE PRIZE
Non-open book prize: £25 prize for the best result (without the use of other sources of information)
Open book prize : £25 prize for the best result with use of other sources of information

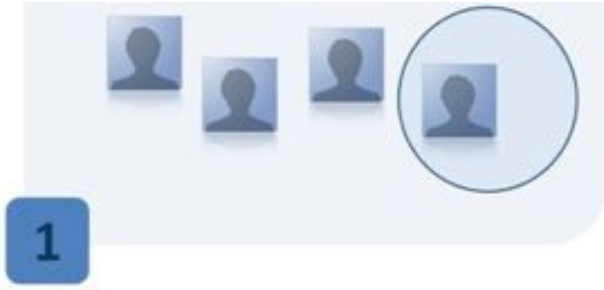
"I already answer the Clinical Reasoning Test what should I do?" If you have already answer the Clinical Reasoning Test, during its' research phase, you can either close touchstone window now and do not complete this test or complete the test now in the final version and get feedback on your results.Plesae choose one of the following options:

I want to take the test again and receive feedback on my new results. ☐
I did not took any of the CRT cases before ☐

1. Agreement and Feedback

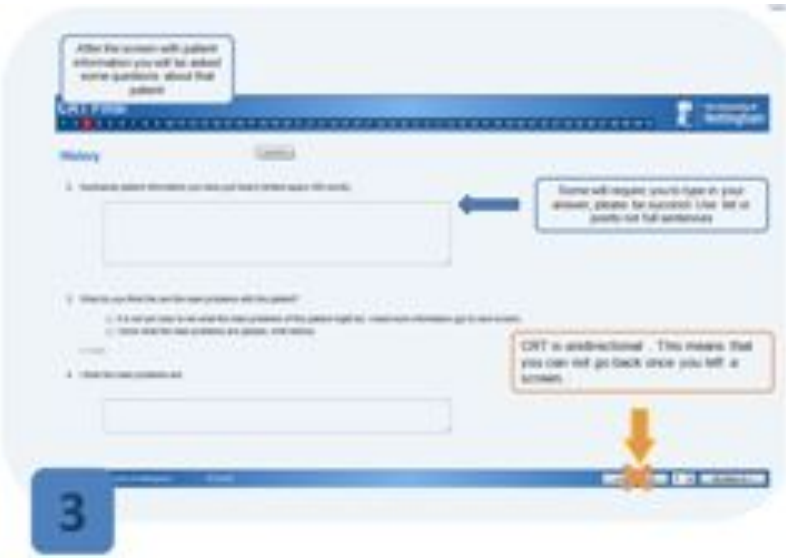
	Yes	No
Do you agree in taking part in this study ?	☐	☐
Do you wish to receive feedback on your results ?	☐	☐

2. Please select from the option below your medical school (pre-clinical years).
Graduate Entry Medical School Derby ☐
Medical School Nottingham ☐
Coimbra Medical School ☐



This test is based on a clinical cases.

First screen will provide you some information about the patient and the context.



You will be asked a sequence of questions about the case.
This test is **unidirectional,you cannot go back once you've left a screen**!

Mr Kravinec

Where do you meet this patient? A&E department
Who are you? Casualty officer

- Forename: Bruno
- Surname: Kravinec
- Sex: male
- Age at Presentation: 60
- Nationality: Polish living in England for 30 years

Mr. Bruno was brought to A&E by ambulance. He appears to be very agitated and confused. He looks pale and his hands are very sweaty. His English is not perfect he seams to be having difficulties explaining his symptoms due to the pain and shortness of breath.He tells you that his boss called the ambulance and that his wife, that is still at work, has been informed he came to hospital.

Listen to his complaints:

History

3. Summarise the patient information you have just heard (limited space 300 words).

(10 marks)

4. **Note:** Research Question

Can you list the main clinical problems of this patient

- [ ] I have some ideas about what the clinical problems might be are but I need more information before listing them (skip text box)
- [ ] I'm ready to list the main clinical problems (use text box below to list them)

5. List here the main clinical problems:

(3 marks)

6. If you could choose the course of action to follow, what would you like to do immediately?

- [ ] I don't know
- [ ] Take history
- [ ] Physical examination
- [ ] Order clinical investigations
- [ ] Start treatment

(1 mark)

7. Please explain why would you take that course of action.

(5 marks)

History

8. What questions would you want to ask Mr Kravinec?
Select 3 for each of the categories below according to their discriminatory value in this case

i. Essential questions

- [ ] Do you have any allergies?
- [ ] What do you do for a living?
- [ ] Have you ever had a heart attack (in the past)?
- [ ] Have you eaten spicy food recently?
- [ ] What were you doing when the pain started ?
- [ ] Do you have any cough with mucus/sputum recently
- [ ] Have you had stomach problems, reflux or heartburn recently?
- [ ] Do you have any pain in your jaw/arm?
- [ ] Have you lost any weight recently?

ii. Important but not very discriminatory

- [ ] Do you have any allergies?
- [ ] What do you do for a living?
- [ ] Have you ever had a heart attack (in the past)?
- [ ] Have you eaten spicy food recently?
- [ ] What were you doing when the pain started ?
- [ ] Do you have any cough with mucus/sputum recently
- [ ] Have you had stomach problems, reflux or heartburn recently?
- [ ] Do you have any pain in your jaw/arm?
- [ ] Have you lost any weight recently?

iii. Least relevant questions

- [ ] Do you have any allergies?
- [ ] What do you do for a living?
- [ ] Have you ever had a heart attack (in the past)?
- [ ] Have you eaten spicy food recently?
- [ ] What were you doing when the pain started ?
- [ ] Do you have any cough with mucus/sputum recently
- [ ] Have you had stomach problems, reflux or heartburn recently?
- [ ] Do you have any pain in your jaw/arm?
- [ ] Have you lost any weight recently?

(9 marks)

History

These are Mr Kravinec answers' to the previous questions

- Do you have any allergies? No
- What do you do for a living? I'm a carpenter
- Have you ever had a heart attack (in the past)? No
- Have you eaten spicy food recently? No
- What were you doing when the pain started? I was working on the frameworks of this old house
- Do you have a any cough with mucus/sputum recently? No
- Have you had stomach problems, reflux or heartburn recently? No, I have some indigestion occasionally
- Do you have any pain in your jaw/arm? No
- Have you lost any weight recently? No

9. If you would like to ask any other questions to the patient, please write them here:

(5 marks)

10. **Note:** Research Question

What do you think are possible differential diagnosis in this case?

- [ ] It is not yet clear to me, I need more information (go to next screen)
- [ ] I have a clear idea of the patient differential diagnosis possibilities (please, write below)

11. Use the table below to write your differential diagnosis at this moment. Please note you do not need to fill in all the spaces available, use only the ones you think necessary.

Seriousness	Likelihood	
	High	Low
High		
Low		

(8 marks)

Physical Examination

12. **Note:** Research Question

Do you know what results to expect from the physical examination ?

- [ ] I have a vague idea of what I may find but I prefer to wait for the results (go to next screen)
- [ ] Yes

13. Select from the list 3 results you would expect to find in the physical examination?

- [ ] Upper abdominal rebound tenderness
- [ ] Splenomegaly
- [ ] Absent bowel sounds
- [ ] Flanks dull to percussion
- [ ] Tachycardia/Tachypnoea
- [ ] Hepatomegaly
- Pulsatile expansile mass

(3 marks)

Physical Examination

On physical examination you find the following:

- BMI:20
- Heart rate: 110bpm
- BP: 110/60mmHg
- Respiratory Rate: 28 per minute
- Chest examination: normal
- Abdominal examination: normal (entirely non tender)

14. **Note:** Research Question

    Have these findings changed your differential diagnosis?

    Yes (write your new diagnosis in the matrix)

    No (go to next screen)

15. Use the table below to write your differential diagnosis. Please note you do not need to fill in all the spaces available, use only the ones you think necessary.

Seriousness	Likelihood	
	High	Low
High		
Low		

(8 marks)

Clinical Investigations

16. Consider the following table

	Name of clinical invest.	Tests included
A	Blood tests (Full Blood Count)	Haemoglobin (Hb); Red Blood (Cell) Count (RBC); Haematocrit (HCT); Mean Cell Volume (MCV) cell haemoglobin (MCH); Mean cell haemoglobin concentration (MCHC); White Blood (Cell) Cou Platelets (PLT); Neutrophils (NEUTS); Lymphocytes (LYMPH); Monocytes (MONOS);Eosinophils (EOS);Basophils (BASOS);Erythrocyte sedimentation rate (ESR);Reticulocytes
B	Urea and Electrolytes	Creatinine (CREA) ; Glucose (GLU) [Plasma] Potassium (K) ; Sodium (NA) ;Urea (UREA)
C	Blood Sample (bioq.)	Albumin (ALBG); Alkaline Phosphatase (ALP) ; Alanine Transaminase (ALT); Aspartate Transam (AST) ; Bicarbonate (BIC) HCO3Bilirubin (BILI) ; Calcium (CAL); Calcium Corrected (CALA);Chl Creatinine (CREA); Gamma-Glutamyl Transpeptidase (GGT); Glucose (GLU) [Plasma]; Osmolal (OSMS); Potassium (K); Protein, Total Serum (TP); Sodium (NA); Urea (UREA).
D	Arterial Blood Gas (ABG)	Bicarbonate (BIC) HCO3; pH; PCO2; PO2; Base Excess
E	Capillary glucose	----------
F	Troponin and CK-MB (cardiac markers)	Creatine Kinase (CK); Creatine Kinase (CK-MB) testLactate Dehydrogenase (LDH); Myoglobin ([Urine, Random]Troponin T (TnT) [Serum]
G	Urinalysis	Blood; Protein; Glucose Ketones; Nitrates; White cells
H	Stool Culture	----------
I	Chest X-ray	----------
J	Abdominal X-ray	----------
K	Abdominal CT scan	----------
L	Coronary Angiogram	----------
M	Upper abdominal Endoscopy	----------
N	ECG	----------

Choose 5 clinical investigations you would like to order now:

- [ ] A
- [ ] B
- [ ] C
- [ ] D
- [ ] E
- [ ] F
- [ ] G
- [ ] I
- [ ] J
- [ ] K
- [ ] L
- [ ] M
- [ ] N

17. Write here the results you expect to find in the clinical investigations

(2 marks)

Choose how will each of the following facts affect the likelihood of your main differential diagnosis?
(will it become more likely? or unlikely?)

	Extremely Unlikely	Not likely	Will not affect the hypotheses	Likely	Highly likely
18. ...Raised troponin and CK-MB	☐	☐	☐	☐	☐
19. ...Chest-X ray normal	☐	☐	☐	☐	☐
20. …Faecal blood test positive for blood	☐	☐	☐	☐	☐
21. ...High O2 and low PaCO2	☐	☐	☐	☐	☐

(5 marks)

22.	...ECG shows a raised ST segment	☐	☐	☐	☐	☐
23.	...C-reactive protein (hs-CRP test) is higher than 2.4 mg/l (normal range: 0-1.0 mg/l)	☐	☐	☐	☐	☐

Clinical Investigations: Results

These are the results from Mr Bruno Kravinec's clinical investigations.

List of investigations **not preformed**:

- Blood Sample (bioq.)
- Capillary glucose
- Urinalysis
- Stool Culture
- Abdominal X-ray
- Abdominal CT scan
- Coronary Angiogram
- Upper abdominal Endoscopy

NHS Trent Region

FULL BLOOD COUNT

Name	Value	Units	Normal Range
Haemoglobin (Hb)	15.7	g/dL	(male) 13.0 - 18.0 g/dL (female) 11.5 - 16.0 g/dL Sex difference negligible until adulthood. (children) 2-6 yrs 13.5 - 18 g/dL
Red Blood (Cell) Count (RBC)	4.2	$x10^{12}/L$	(male) 4.5 - 6.5 $x10^{12}/L$ (female) 3.8 - 5.8 $x10^{12}/L$
Haematocrit (HCT)	43.4%		(male) 0.38 - 0.52 (38% - 52%) (female) 0.35 - 0.47 (35% - 47%)
Mean Cell Volume (MCV)	96.2	fL	80 - 98 fL Cells are larger in neonates, though smaller in other children.
Mean cell haemoglobin (MCH)	26*	pg	26 - 34 pg
Mean cell haemoglobin concentration (MCHC)	28*	g/dL	32 - 36 g/dL
White Blood (Cell) Count (WBC)	9.8	$x10^9/L$	4.0 - 11.0 $x10^9/L$ Higher in neonates and infants
Platelets (PLT)	327	$x10^9/L$	150 - 400 $x10^9/L$
Neutrophils (NEUTS)	2.3	$x10^9/L$	2.0 - 7.5 $x10^9/L$ (pediatric: 1.0 - 9.0 $x10^9/L$)
Lymphocytes (LYMPH)	3.7	$x10^9/L$	1.5 - 4.0 $x10^9/L$ (pediatric: 2.5 - 16.5 $x10^9/L$)
Monocytes (MONOS)	0.25	$x10^9/L$	0.2 - 0.8 $x10^9/L$ (pediatric: mean 0.7 $x10^9/L$)
Eosinophils (EOS)	0.02	$x10^9/L$	0.02 - 0.5 $x10^9/L$ (pediatric: 0.0 - 0.3 $x10^9/L$)
Basophils (BASOS)	0.2	$x10^9/L$	0.0 - 0.2 $x10^9/L$
Erythrocyte sedimentation rate (ESR)	29	mm/hr	Female: <20 mm/hr Male: <10 mm/hr ESR increases with age.
Reticulocytes	-	$x10^9/L$	10 - 100 $x10^9/L$

NHS Trent Region

Urea and Electrolytes - Taken on admission

Name	Value	Units	Normal Range
Creatinine (CREA) [Serum]	78	µmol/L	Women 55 - 100 µmol/L, Men 70 - 120 µmol/L Muscle mass effects results: the lower the mass the lower the creatinine.
Glucose (GLU) [Plasma]	6.0	mmol/L	Fasting 3.0 - 6.0 mmol/L
Potassium (K) [Serum]	3.0	mmol/L	3.5 - 5.3 mmol/L
Sodium (NA) [Serum]	140	mmol/L	134 -145 mmol/L
Urea (UREA) [Serum]	3.8	mmol/L	Adults Less than 60 yrs 2.0 - 6.5 mmol/L Greater than 60yrs 2.9 - 7.5 mmol/L

Arterial Blood Gas - Taken on admission

Name	Value	Units	Normal Range
Bicarbonate (BIC) HCO_3	24	mmol/L	22 - 26 mmol/L
PCO_2	5.0	kPa	4.7 - 6.0 kPa
pH	7.33		7.34 - 7.44 (H+ ion 36-42 nmol/L)
PO_2	11.4	kPa	11.3 - 14.0 kPa
Base Excess	-1		Base Excess -2.3 - +2.3

NHS Trent Region

Troponin and CK-MB (cardiac markers)

Name	Value	Units	Normal Range
Creatine Kinase (CK) [Plasma, Heparin]	195*	U/L	Women 35 - 143 U/L, Men Up to 35 - 194 U/L. CK results will be higher (up to 2x the upper limit of the [illegible] and in patients of Afro-Caribbean origin
Creatine Kinase (CK-MB) test	3.34	ng/mL	0-3 ng/mL
[illegible] [Plasma]	[illegible]	[illegible]	[illegible]
Myoglobin (MYO) [Urine, Random]	N/A	µg/l	N/A
Troponin T (TnT) [Serum]	0.09	µg/L	< 0.03 µg/L

Diagnosis

24. Based on the entire history, examination and investigations, what do you consider is now the most likely diagnosis?

(5 marks)

25. Why do you think that is the most likely diagnosis?(Provide reasons for your answer to the previous question)

(5 marks)

NOTE: Research Question

26. How confident are you that the diagnosis you have indicated is correct?

	Not confident at all	Not confident	I don\'t know	Confient	Extremely confident
	☐	☐	☐	☐	☐

NOTE: Research Questions

Rate how each part of the case contributed toward your decision

	Irrelevant	Not important	I don\'t know	Important	Very important
27. History	☐	☐	☐	☐	☐
28. Physical examination results	☐	☐	☐	☐	☐
29. Clinical investigations	☐	☐	☐	☐	☐

Chest X-ray

Chest X-ray	Normal

ECG

ECG	4mm of ST elevation V2-V4 with ST depression leads to II to III and AVL

30. **Clinical Vignette**

Mrs Kravinec comes in to check on her husband. She is highly concerned because he´s been having health problems. She also provides you with new information about this patient.

Hypothesis	New Information
Myocardial Infarction (STEMI)	> Mr Kravinec was a heavy smoker until 3 years ago, when he completely stop smoking. > He´s had NSAID induced peptic ulcer. > He´s also had a deep vein thrombosis 6 months ago. > He is taking warfarin.

Then this hypothesis is:

- ○ ruled out or almost ruled out
- ○ less probable
- ○ neither less or more probable
- ○ more probable
- ○ certain or almost certain

(1 mark)

Mr Kravinec: Final questions

31. **Note:** Research Question

Briefly explain the pathophysiological mechanism of Myocardial Infarction?

(1 mark)

32. **Note:** Research Question

Have you ever encountered a similar case?

- [ ] No, it was absolutely new to me
- [ ] No, but I have had contact with cases with similar features before
- [ ] Yes

33. **Note:** Research Question

If appropriate please indicate where you have encountered a similar case

- [ ] I have not encountered anything like this case
- [ ] Lectures
- [ ] PBL scenarios
- [ ] Clinical rotations

Mr Kravinec: Final questions

34. **Note:** Research Question

Did you look up any of the answers on books or use any other source of information to answer this case? (you can choose more then one option)

- [ ] No, I did not used additional source of information
- [ ] Yes, text books
- [ ] Yes, online resources such as WebMD, BMJ or others
- [ ] Yes, course materials from NLE

(1 mark)

NOTE: Research Question

	N/A	Not useful at all (I could not find any relevant information at all)	Not very useful (It help me t answer only a few questions, not all)	Not Applicable	Quite useful (It provided me with answer to most of the questions)	Extremely useful (I found all answer in the text book)
35. If you used any additional source of information, please rate how useful you think the information was in order to provide correct answers to this test	☐	☐	☐	☐	☐	☐

SUDOKO EXERCISE

THIS IS OPTIONAL AND IT WILL NOT BE MARKED, so please ignore the marks underneath.

A sudoku puzzle is a grid 9 squares wide and 9 squares deep. The lines of squares running horizontally are called rows, and the lines running vertically are called columns.

The grid is further divided by the darker lines into nine 3 X 3 square 'boxes'.Some of the squares already have numbers in them. Your task is to fill in the blank squares.

There's only one rule: **Each row, column and box must end up containing all of the numbers from 1 to 9.**

This rule has an important side-effect, which is the basis of all solving techniques:Each number can only appear once in a row, column or box (source: RandomSudoku@paulspages.co.uk).

36. Sudoko Exercise

6		1				3		4
		5	1		9	8		
	8						7	
1			2	9	5			7
				1				
2			6	3	4			8
	1						5	
		6	7		2	4		
8		7				9		

(51 marks)

Mr Bruno Kravinec

References:

- White, H.D. & Chew, D.P., 2008. Acute myocardial infarction. Lancet, 372(9638), pp.570-84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18707987
- NHS. NHS Clinical Knowledge Summaries. National Excellence, Institute For Health And Clinical. 2010. Available at: http://www.cks.nhs.uk/home. NHS. Map of Medicine. Institute for Innovation and Improvement. 2010. Available at: http://www.mapofmedicine.com/
- Trialists A, Collaboration ATT. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. The Lancet. 2009;373(9678):1849-1860. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0140673609605031.
- Macdonald JE, Struthers AD, B MBCH, C BS. What Is the Optimal Serum Potassium Level in Cardiovascular Patients ? Journal of the American College of Cardiology. 2010.
- Road H. Cardiac arrhythmias following acute myocardial infarction : Associations with the serum potassium level and prior diuretic therapy. European Heart Journal. 1984:464-469.

Miss Karen Smith

Where do you meet this patient? A&E department

Who are you? Casualty officer

- Forename: Karen
- Surname: Smith
- Sex: female
- Age at Presentation: 17
- Nationality: English

Miss Smith comes to with her mother. She is a petite girl with fair skin and blond hair. She looks very pale and weak. Her skin looks dry with some smalls cracks and she has very dry lips . She seams to be dizzy, having difficulty in concentrating and telling you how she is feeling. She speaks quietly making it difficult for you to hear correctly what she is saying.

Listen to her complaints:

History

37. Summarize the patient information you have just read and heard (limited space 300 words).

(1 mark)

38. **Note:** Research Question

Can you list the main clinical problems of this patient

- [ ] I have some ideas about what the clinical problems might be are but I need more information before listing them (skip text box)
- [ ] I\'m ready to list the main clinical problems (use text box below to list them)

39. List here the main clinical problems:

(1 mark)

40. If you could choose the course of action to follow, what would you like to do immediately?

- [ ] I don't know
- [ ] Take history
- [ ] Physical examination
- [ ] Order clinical investigations
- [ ] Start treatment

41. Please explain why would you take that course of action.

(1 mark)

History

These are Miss Smith's answers to the previous questions

- Are you diabetic? Yes and I use insulin
- Has anyone you know had similar symptoms? No
- Do you have any allergies? Penicillin
- Are you taking any medication? Just the contraceptive pill
- Has the pain moved since it started? No
- Have you noticed any swelling in your legs recently? No
- Have you ever been pregnant? No
- What do you do for a living? I'm a student
- When was the start of your last menstrual period? 2 weeks ago
- Do you drink a lot of caffeinated drinks? Yes, some

42. Write here, if you have additional questions you would like to ask the patient

(1 mark)

43. **Note:** Research Question

What do you think are possible differential diagnosis in this case?

- [ ] It is not yet clear to me, I need more information (go to next screen)
- [ ] I have a clear idea of the patient differential diagnosis possibilities (please, write below)

44. Use the table below to write your differential diagnosis. Please note you do not need to fill in all the spaces available, use only the ones you think necessary.

Seriousness	Likelihood	
	High	Low
High		
Low		

(8 marks)

Physical Examination

45. **Note:** Research Question

Do you know what results to expect from the physical examination ?

- [ ] I have a vague idea of what I may find but I prefer to wait for the results (go to next screen)
- [ ] Yes

46. Select 3 results you would expect to find in the physical examination?

- [ ] Dry mucous membranes
- [ ] Pyrexia
- [ ] Raised JVP
- [ ] Strong smelling breath
- [ ] Right iliac fosse rebound tenderness
- [ ] Absent bowel sounds
- [ ] Tachycardia/Tachypnoa

(3 marks)

Physical Examination

On physical examination you find the following:

- BMI:20
- Heart rate: 130bpm
- BP: 80/50mmHg
- Respiratory Rate: 34 per minute
- Chest examination: normal
- Abdominal examination: mild non specific generalized tenderness
- Patient looks very pale with clear signs of dehydration and confusion

47. **Note:** C

Have these findings changed your differential diagnosis?

- [ ] Yes
- [ ] No (go to next screen)

(1 mark)

48. Use the table below to write your differential diagnosis. Please note you do not need to fill in all the spaces available, use only the ones you think necessary.

Seriousness	Likelihood	
	High	Low
High		
Low		

(8 marks)

Clinical Investigations

49.

	Name of clinical invest.	Tests included
A	Blood tests (Full Blood Count)	Haemoglobin (Hb); Red Blood (Cell) Count (RBC); Haematocrit (HCT); Mean Cell Volume (MCV) cell haemoglobin (MCH); Mean cell haemoglobin concentration (MCHC); White Blood (Cell) Cou Platelets (PLT); Neutrophils (NEUTS); Lymphocytes (LYMPH); Monocytes (MONOS);Eosinophils (EOS);Basophils (BASOS);Erythrocyte sedimentation rate (ESR);Reticulocytes
B	Urea and Electrolytes	Creatinine (CREA) ; Glucose (GLU) [Plasma] Potassium (K) ; Sodium (NA) ;Urea (UREA)
C	Blood Sample (bioq.)	Albumin (ALBG); Alkaline Phosphatase (ALP) ; Alanine Transaminase (ALT); Aspartate Transam (AST) ; Bicarbonate (BIC) HCO3Bilirubin (BILI) ; Calcium (CAL); Calcium Corrected (CALA);Chl Creatinine (CREA); Gamma-Glutamyl Transpeptidase (GGT); Glucose (GLU) [Plasma]; Osmolal (OSMS); Potassium (K); Protein, Total Serum (TP); Sodium (NA); Urea (UREA).
D	Arterial Blood Gas (ABG)	Bicarbonate (BIC) HCO3; pH; PCO2; PO2; Base Excess
E	Capillary glucose	----------
F	Troponin and CK-MB (cardiac markers)	Creatine Kinase (CK); Creatine Kinase (CK-MB) testLactate Dehydrogenase (LDH); Myoglobin ([Urine, Random]Troponin T (TnT) [Serum]
G	Urinalysis	Blood; Protein; Glucose Ketones; Nitrates; White cells
H	Stool Culture	----------
I	Chest X-ray	----------
J	Abdominal X-ray	----------
K	Abdominal CT scan	----------
L	Coronary Angiogram	----------
M	Upper abdominal Endoscopy	----------
N	ECG	----------

Chose from the following table the 5 investigations you would like to order first.

- [ ] A
- [ ] B
- [ ] C
- [ ] D
- [ ] E
- [ ] F
- [ ] G
- [ ] H
- [ ] I
- [ ] J
- [ ] K
- [ ] L
- [ ] M
- [ ] N

50. Write here 2 results you expect to find in the clinical investigations

(6 marks)

How will each of the following facts affect the probability of your main differential diagnosis? *(will it*

become more likely? or unlikely?)

	Extremely Unlikely	Not likely	Will not affect the hypotheses	Likely	Highly likely
51. ...Urinalysis reveals normal levels of glucose and ketones	☐	☐	☐	☐	☐
52. ...Chest X-ray is normal	☐	☐	☐	☐	☐
53. ...Slight increase in Hb (17g/dl)	☐	☐	☐	☐	☐
54. ...Serum amylase is normal	☐	☐	☐	☐	☐

Clinical Investigations: Results

These are the results from Miss Smith's clinical investigations.

List of investigations **not preformed**:

- Blood Sample (bioq.)
- Troponin and CK-MB
- Stool Culture
- Chest X-ray
- Abdominal X-ray
- Abdominal CT scan
- Coronary Angiogram
- Upper abdominal Endoscopy
- ECG

NHS Trent Region

FULL BLOOD COUNT

Name	Value	Units	Normal Range
Haemoglobin (Hb)	17	g/dL	(male) 13.0 - 18.0 g/dL (female) 11.5 - 16.0 g/dL Sex difference negligible until adulthood. (children) 2-6 yrs 13.5 - 18 g/dL
Red Blood (Cell) Count (RBC)	5.8	$\times 10^{12}$/L	(male) 4.5 - 6.5 $\times 10^{12}$/L (female) 3.8 - 5.8 $\times 10^{12}$/L
Haematocrit (HCT)	0.58		(male) 0.38 - 0.52 (38% - 52%) (female) 0.35 - 0.47 (35% - 47%)
Mean Cell Volume (MCV)	90	fL	80 - 98 fL Cells are larger in neonates, though smaller in other children.
Mean cell haemoglobin (MCH)	30	pg	26 - 34 pg
Mean cell haemoglobin concentration (MCHC)	29	g/dL	32 - 36 g/dL
White Blood (Cell) Count (WBC)	13.1	$\times 10^{9}$/L	4.0 - 11.0 $\times 10^{9}$/L Higher in neonates and infants
Platelets (PLT)	390	$\times 10^{9}$/L	150 - 400 $\times 10^{9}$/L
Neutrophils (NEUTS)	8.3	$\times 10^{9}$/L	2.0 - 7.5 $\times 10^{9}$/L (pediatric: 1.0 - 9.0 $\times 10^{9}$/L)
Lymphocytes (LYMPH)	3.7	$\times 10^{9}$/L	1.5 - 4.0 $\times 10^{9}$/L (pediatric: 2.5 - 16.5 $\times 10^{9}$/L)
Monocytes (MONOS)	0.67	$\times 10^{9}$/L	0.2 - 0.8 $\times 10^{9}$/L (pediatric: mean 0.7 $\times 10^{9}$/L)
Eosinophils (EOS)	0.36	$\times 10^{9}$/L	0.02 - 0.5 $\times 10^{9}$/L (pediatric: 0.0 - 0.3 $\times 10^{9}$/L)
Basophils (BASOS)	0.14	$\times 10^{9}$/L	0.0 - 0.2 $\times 10^{9}$/L
Erythrocyte sedimentation rate (ESR)	29	mm/hr	Female: <20 mm/hr Male: <10 mm/hr ESR increases with age.

NHS
Trent Region

UREA AND ELECTROLYTES - TAKEN ON ADMISSION

Name	Value	Units	Normal Range
Creatinine (CREA) [Serum]	78	µmol/L	Women 55 - 100 µmol/L; Men 70 - 120 µmol/L Muscle mass effects results: the lower the mass the lower the creatinine. Paediatric reference ranges: Child 27 - 62 µmol/L
Glucose (GLU) [Plasma]	6.0	mmol/L	Fasting 3.0 - 6.0 mmol/L
Potassium (K) [Serum]	4.9	mmol/L	3.5 - 5.3 mmol/L
Sodium (NA) [Serum]	152	mmol/L	134 -145 mmol/L
Urea (UREA) [Serum]	12	mmol/L	Adults Less than 60 yrs 2.0 - 6.5 mmol/L Greater than 60yrs 2.9 - 7.5 mmol/L

NHS
Trent Region

ARTERIAL BLOOD GASSES - TAKEN ON ADMISSION

NAME	VALUE	UNITS	NORMAL RANGE
Bicarbonate (BIC) HCO_3	12	mmol/L	22 - 26 mmol/L
PCO_2	1.7	kPa	4.7 - 6.0 kPa
pH	7		7.34 - 7.44 (H+ ion 36-42 nmol/L)
PO_2	13.6	kPa	11.3 - 14.0 kPa
Base Excess	-2.8		Base Excess -2.3 - +2.3

NHS
Trent Region

URINALYSIS

Urinalysis	Blood: None Protein: None Glucose +++ Ketones: +++ Nitrates: None White cells: None

CAPILARY GLUCOSE

Name	Value	Units	Normal Range
Capillary glucose	24	mmol/L	3.6 – 5.8 mmol/L

Diagnosis

55. Based on the entire history, examination and investigations, what do you consider is now the most likely diagnosis?

(1 mark)

56. Why do you think that is the most likely diagnosis?(Provide reasons for your answer to the previous question)

(1 mark)

NOTE: Research Question

Not confident at all | Not confident | I don\'t know | Confident | Extremely confident

57. How confident are you that the diagnosis you have indicated is correct?

Rate how each part of the case contributed toward your decision.

	Irrelevant	Not important	I don\'t know	Important	Very important
58. History					
59. Physical examination results					
60. Clinical investigations					

Miss Karen Smith: Final questions

61. **Note:** Research Question

Briefly explain the pathophysiological mechanism of DKA?

(1 mark)

62. **Note:** Research Question

Have you ever encountered a similar case?

- No, it was absolutely new to me
- No, but I have had contact with cases with similar features before
- Yes

63. **Note:** Research Question

If appropriate please indicate where you have encountered a similar case

(1 mark)

Mr Kravinec: Final questions

64. **Note:** Research Question

Did you look up any of the answers on books or use any other source of information to answer this case? (you can choose more then one option)

- No, I did not used additional source of information
- Yes, text books
- Yes, online resources such as WebMD, BMJ or others
- Yes, course materials from NLE

(1 mark)

NOTE: Research Question

Not useful at all (I Not very useful (It Quite useful (It Extremely

	N/A	could not find any relevant information at all)	help me t answer only a few questions, not all)	Not Applicable	provided me with answer to most of the questions)	useful (I found all answer in the text book)
65. If you used any additional source of information, please rate how useful you think the information was in order to provide correct answers to this test	☐	☐	☐	☐	☐	☐

References

- Hamy, Osama. (2009) Diabetic Ketoacidosis, Harvard Medical School Contributor Information and Disclosures.
- Joint British Diabetes Societies Inpatient Care Group: The Management of Diabetic Ketoacidosis in Adults - March 2010 available at: http://www.diabetes.nhs.uk/publications_and_resources/reports_and_guidance/
- NHS. NHS Clinical Knowledge Summaries. National Excellence, Institute For Health And Clinical. 2010. Available at: http://www.cks.nhs.uk/home.NHS.
- •Map of Medicine. Institute for Innovation and Improvement. 2010. Available at: http://www.mapofmedicine.com/
- Kitabchi AE, Nyenwe EA. Hyperglycemic crises in diabetes mellitus: diabetic ketoacidosis and hyperglycemic hyperosmolar state. Endocrinol Metab Clin North Am. 2006,35:725-751.
- Wachtel TJ, Tetu-Mouradjian LM, Goldman DL, et al. Hyperosmolarity and acidosis in diabetes mellitus: a three-year experience in Rhode Island. J Gen Intern Med. 1991;6:495-502.
- Joint British Diabetes Societies Inpatient Care Group. The Management of Diabetic Ketoacidosis in Adults. March 2010. http://www.diabetes.nhs.uk/ (last accessed 06 July 2010).

Was this useful?

NOTE: Research Question

	1	2	3	4	5
66. On Scale of 1(unuseful) to 5 (extremely useful) please rate how useful do you think the CRT type of tests could be for preparing you to clinical practice.	☐	☐	☐	☐	☐

67. **Note:** Research Question

Would you think online cases, such as the ones you answered and feedback, would help develop your clinical/diagnostic reasoning skills?

- ☐ yes
- ☐ no
- ☐ Other

68. **Note:** Research Question

Would you like to have access to more cases like these for revision and study purposes?

- ☐ no
- ☐ yes

Thank you!

THANK YOU SO MUCH FOR TAKING PART IN THIS STUDY!

You will shortly receive your feedback and the certificate of participation in Medical Education Research.
Results of the prize will be announced as soon as all data is collected.

Kind Regards