Evaluation of an online learning tool to improve medical students' clinical reasoning skills

Ruth Louise Plackett

UCL

A thesis submitted for the degree of Doctor of Philosophy

March 2019

Declaration

I, Ruth Plackett confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

Statement of contribution

This PhD is part of a project funded by the NIHR Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis (PRU). A previous PhD funded by the PRU and supervised by Raine and Sheringham, found that GPs were influenced by cognitive biases when making clinical decisions using an online patient vignette tool (Sheringham et al., 2016). The PRU provided further funding to adapt this online patient vignette tool into a new training tool to help reduce the impact of cognitive biases and to evaluate this tool using a trial. I was responsible for ensuring the tool was developed based on current theory and evidence and the evaluation of the tool.

I used my findings from my literature review, theoretical model, systematic literature review and from user testing to contribute towards the development of an online patient simulation learning tool. As part of the eCREST team I also played a role in many decisions concerned with the tool's content and functionality.

I screened, extracted and assessed the quality of the articles included in my systematic review with the assistance of APK and SM, who screened, extracted quality assessed a proportion of the articles. I analysed all the data from my systematic review. A group of four GP Registrars, NK, SM, JH and PS developed the clinical content for the patient cases. A web developer, Silver District, developed the website. My supervisor (APK) was the project manager and oversaw the development of the tool.

I conducted the evaluation of the tool. I was required to conduct a trial to evaluate the simulation as stipulated by the PRU but chose the type of trial, outcome measures and qualitative research methods to evaluate eCREST. I collected data for the trial and Think Aloud and interviews. I was assisted in data collection of the Think Aloud and interviews by three medical students on the Peer Assisted Learning scheme (PALs). A group of registrars SM, JH, PS, SG and JT contributed to the development of a measure of clinical reasoning skills. I analysed all the data from my feasibility RCT and Think Aloud study. JT, APK and MK contributed to the data analysis of the Think Aloud and interview study by reading the transcripts and double coding a proportion of the transcripts. When referring to the research team in this report I am referring to the GP registrars, my supervisors and myself.

Abstract

Background: Evidence suggests that problems in clinical reasoning skills – the thought processes required to make clinical decisions – are the leading cause of diagnostic errors, which can lead to significant patient harm. Theories of learning and clinical reasoning have indicated online patient simulations (OPS) could be a novel approach to improving medical students' clinical reasoning skills. However, little is known about their impact on clinical reasoning.

Methods: I conducted a systematic literature review to explore the effectiveness of OPS. Informed by my review and theory, I co-developed eCREST (electronic Clinical Reasoning Skills Educational Simulation Tool). I assessed the feasibility, acceptability and potential impact of eCREST at three UK medical schools with a feasibility randomised controlled trial (RCT). I explored how students reasoned when using eCREST and what factors influenced reasoning, using a Think Aloud and interview approach with 16 medical students.

Results: My systematic review found OPS may be effective at improving medical students' clinical reasoning skills but the few studies available lacked methodological rigour, so these results should be treated with caution. Uptake and retention in the feasibility trial was acceptable and provided evidence to support a definitive RCT. Impact data suggested eCREST may improve clinical reasoning skills - the intervention group were significantly more likely to gather essential information from the 'patient' than controls ($OR = 1.4$; 95% CI 1.1-1.7, $n = 148$). Qualitative findings suggested that students use a variety of data gathering strategies

and eCREST helped students to structure their data gathering and stay open-minded about diagnosis. Students' knowledge, confidence and engagement with eCREST also influenced these strategies.

Conclusions: Tools like eCREST can improve reasoning skills by helping students to gather essential information and potentially reduce future missed diagnostic opportunities. Evaluations of such tools are now needed within medical curricula, using validated outcome measures to determine effectiveness.

Impact Statement

My research has advanced the current understanding of how online patient simulations can improve clinical reasoning skills in a number of ways. Through the development of an online patient simulation logic model, grounded in a wider theory of clinical reasoning amongst medical students, I have set out important targets for such tools, possible mechanisms of change and strategies for measuring valid and sensitive outcomes. The results from my feasibility randomised controlled trial (RCT) showed promising indications that online patient simulations can improve clinical reasoning skills but a definitive RCT is needed to establish effectiveness. My PhD has also provided insights into how medical students reason and how simulations can affect the different ways students reason, through in-depth qualitative data. Impact from these findings will be demonstrated through peer-reviewed publications of the empirical chapters of my PhD.

My research may also have implications for how clinical reasoning is taught in medical schools. It has provided proof of concept for the implementation of online patient simulations into medical curricula to teach clinical reasoning skills. Through dissemination of my findings at national conferences, there has been interest in implementing eCREST in other UK medical schools that did not participate in my feasibility trial. Additionally, those medical schools who took part in my trial have shown willingness to continue using eCREST and agreed to discuss integrating it into their curricula. This demonstrates that there is appetite for the use

of novel technologies to teach clinical skills at medical schools. There has also been international interest, with one site in Turkey having already used eCREST as part of their curriculum.

Given the interest in using eCREST from medical schools, my supervisors and I have joined a research project named EDUCATE to explore how we can provide eCREST to more medical schools. EDUCATE is a six-month training and development programme for educational technology groups, within the UCL Institute of Education, and is part-funded by the European Regional Development Fund. It will provide us with business knowledge and best practice in educational research. Being part of this project will help to inform the strategy for expanding and sustaining the use of eCREST in medical schools in the UK and abroad.

My research may also have an impact on clinical practice and policies to improve healthcare. eCREST aims to train future clinicians how to reason in complex situations in primary care, in which patients present with common but potentially life-threatening conditions that are often missed, such as lung cancer. National policies in the UK and worldwide have emphasised the urgent need to improve the early detection of cancer in primary care. Explicitly teaching medical students clinical reasoning skills in primary care, using online patient simulations, could help students to develop better reasoning skills that they can apply to their clinical work and potentially reduce future missed diagnostic opportunities and patient harm.

Table of contents

List of Tables

List of Figures

Glossary of key terms and list of abbreviations

Key terms

Anchoring: the tendency to stick to an initial hypothesis despite new contradictory information.

Clinical reasoning skills: the thought processes required to identify likely diagnoses, formulate appropriate questions and reach clinical decisions.

Cognitive biases: inherent errors in thinking that deviate from rational thinking.

Confirmation bias: the tendency to seek information to confirm a hypothesis rather than refute it.

Consultation outcomes: the products of the clinical decisions made as part of the clinical reasoning process, such as the diagnosis and management plan for a patient.

Diagnostic errors: the failure to (a) establish an accurate and timely explanation of a patient's health problem(s) or (b) communicate that explanation to the patient.

Diagnostic Thinking Inventory (DTI): a self-reported measure of clinical reasoning skills that consists of two sub-scales; one measures flexibility of thinking, the other measures the structure of knowledge in memory.

Differential diagnosis: the process of differentiating between two or more conditions that share similar signs or symptoms. It commonly results in a list of possible conditions that could be causing symptoms based on the information gathered from a patient.

eLearning (online learning): the use of technology to deliver learning, usually via the internet.

electronic Clinical Reasoning Educational Simulation Tool (eCREST): an online patient simulation learning tool that replicates clinical consultations in general practice. Using reflective prompts and feedback, it improves students' abilities in gathering the necessary information to make an informed diagnosis and keep an open-mind about diagnoses.

Flexibility in Thinking (FIT) sub-scale: sub-scale of the DTI that measures the variety of thought processes clinicians use in the diagnostic process.

General Practitioner (GP): a community-based doctor who treats patients with minor, acute and chronic illnesses and is often the first point of contact in the UK healthcare system for most patients.

GP registrar: a junior doctor who is training to be a GP. They are often based in GP surgeries and are supervised by an approved GP trainer. The GP registrar year is the last year of a 5-year training period to become a GP in the UK.

Key Features Problems: an observed measure of clinical reasoning skills where a clinical problem is presented via a patient case and is typically followed by two or three questions. Questions are designed to assess clinical decisions and relate only to the most important steps in resolving the problem.

Metacognition: the ability to select, monitor and evaluate one's own thinking.

Missed diagnostic opportunities: the retrospective analysis of whether a different action could have been taken by a clinician, which could have led to a correct diagnosis earlier.

MOOCs: Massive Open Online Courses are online courses that are freely available for anyone to enrol.

Online patient simulation (OPS): a specific type of computer-based program that simulates real-life clinical scenarios; learners may emulate the roles of healthcare providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions.

Primary care: the service that provides first point of contact in the healthcare system in the UK and provision of continuity of care. It includes general practice, community pharmacy, dental, and optometry services.

Think Aloud: a qualitative study design that involves observing a participant verbalising their thoughts while completing a task.

Unpacking principle: the tendency not to elicit all the necessary information to make a judgement.

Abbreviations

AAMC: Association of American Medical Colleges.

SB: Urban medical school in the UK that use a system based curriculum and who participated in the feasibility RCT.

COPD: Chronic obstructive pulmonary disease.

DoH: Department of Health.

IoM: Institute of Medicine. Now called the National Academy of Medicine.

IQR: Interquartile range.

M: Mean.

MCQs: Multiple-choice questions.

Mdn: Median.

NHS: National Health Service.

NIHR: National Institute for Health Research.

PALs: Peer Assisted Learning scheme.

PBL: Problem based learning.

PPI: Patient and Public Involvement.

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

PRU: NIHR Policy Research Unit in cancer awareness, screening and early diagnosis.

RAP: Research Advisory Panel.

RCT: Randomised controlled trial.

RUMS: The Royal Free, University College and Middlesex Medical Students' Association.

RQ: Research question.

SD: Standard deviation.

SEM: Standard error.

SRCR: Self-reported clinical reasoning skills.

SA: Urban medical school in the UK that use a system based curriculum and who participated in the feasibility RCT and Think Aloud and interview study.

SC: Rural medical school in the UK that use a PBL based curriculum and who participated in the feasibility RCT.

UK: United Kingdom.

WHO: World Health Organisation.

1 Introduction

Clinical reasoning errors are the leading cause of diagnostic error in medicine and can cause substantial harm to patients (Cheraghi-Sohi et al., 2015; Institute of Medicine [IoM], 2015). There is pressure from policy makers to reduce diagnostic errors and improve patient safety (Department of Health [DoH], 2000; 2015). A relatively unexplored strategy to reduce diagnostic errors is to target the development and teaching of clinical reasoning skills in undergraduate medical education (Higgs & Jones, 2000; IoM, 2015). This thesis describes the development and evaluation of an online patient simulation learning tool that aims to improve the clinical reasoning skills of medical students. This PhD was part of a wider project that aimed to develop educational resources for future doctors to support their diagnostic decision-making.

In this chapter, I outline the key terminology and relevant literature to support the development and evaluation of such a tool. In **Chapter 2**, I describe the aims and objectives for this thesis. I then summarise the current understanding of the effectiveness of online patient simulations designed to improve medical students' clinical reasoning skills through the results of a systematic review (**Chapter 3**). I propose a logic model of how online patient simulations help students learn clinical reasoning skills and the development of a novel learning tool - the electronic clinical reasoning educational simulation tool (eCREST) in **Chapter 4**. In **Chapters 5 and 6**, I describe the feasibility randomised controlled trial (RCT) that was used to assess the feasibility, acceptability and potential impacts of eCREST. This is followed by an account

of a mixed-methods approach that further explores the findings from my trial using a qualitative, Think Aloud and interview study design (**Chapters 7 and 8**). I conclude with a summary of my findings and implications for policy, medical education and research (**Chapter 9**).

1.1 Missed opportunities for diagnosis

Research has estimated that around 10% of patients experience diagnostic error (Graber, 2013; IoM, 2015; Vincent, Neale, & Woloshynowych, 2001). Diagnostic errors can have severe consequences for patient safety (Cheraghi-Sohi et al., 2015). It has been estimated that over 80% of cases of diagnostic error result in some degree of harm to the patient (IoM, 2015; Singh et al., 2013). In the UK, it has been estimated that 5.2% of deaths each year, roughly 12,000 patients, had a 50% or more chance of being preventable, and 29.7% of these deaths were caused by diagnostic error (Hogan et al., 2012). Diagnostic errors can also have other implications. For example, they can have a significant financial impact on the health service. Diagnostic errors often result in higher costs because of unnecessary treatments, lengthy and frequent visits to health services and litigation costs (IoM, 2015; Kostopoulou, Delaney, & Munro, 2008). Indeed, in the UK and US diagnostic error accounts for most litigation claims (Phillips et al., 2004; Silk, 2000). Diagnostic errors can also have an impact on the doctors themselves, as it can affect their reputation and cause personal distress (Fisseni, Pentzek, & Abholz, 2008; West, Huschka, Novotny, & et al., 2006).

1.1.1 Definition of diagnostic errors and missed diagnostic opportunities

A definition of diagnostic error given by The IoM is "the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient" (IoM, 2015). Measuring diagnostic error or preventable death retrospectively is difficult, as researchers are unlikely to have a complete understanding of the information and resources available at that time (Hogan et al., 2012; Singh, 2014). Indeed, some researchers have instead used the term 'missed opportunities' to describe errors (Lyratzopoulos, Vedsted, & Singh, 2015; Singh, 2014). Missed opportunities refer to events/situations where, according to a retrospective analysis, a different action might have been taken by a clinician, which would have led to a correct diagnosis earlier. This missed opportunity could be caused by cognitive or system factors, the cause may be outside of the clinician's control and is framed within the context of an evolving diagnostic process. Not all missed opportunities result in patient harm and not all delays or incorrect diagnoses are a result of missed opportunities (**Figure 1-1**) (Lyratzopoulos et al., 2015; Singh, 2014). For example, atypical presentations of disease can lead to delayed or incorrect diagnoses but there may have been little a clinician could do to reach a correct diagnosis earlier given the presenting symptoms (Singh, 2014). Using the term 'missed diagnostic opportunities' takes the focus away from blaming clinicians for errors and encourages learning from these events (Singh, 2014). Diagnostic errors and missed diagnostic opportunities are both terms that are used in the literature and diagnostic error particularly so in the literature focusing on events attributed to individual clinicians' decisions. In this

thesis, the term missed diagnostic opportunities will be used instead of diagnostic error, as it better captures the system-wide nature of these events.

Figure 1-1 A model defining missed opportunities for diagnosis.
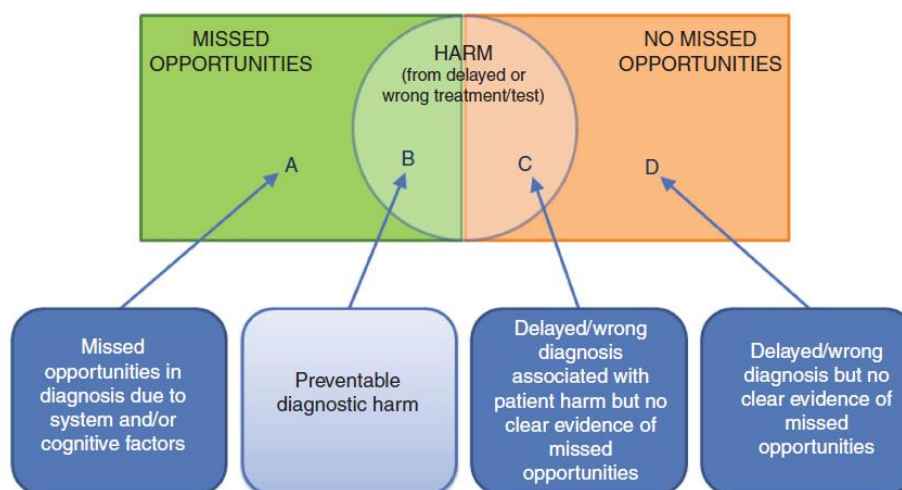
Note: Taken from Lyratzopoulos et al. 2015. Adapted from Singh 2014

1.1.2 Primary care

Primary care is a setting in which missed diagnostic opportunities are likely to be prevalent. This is because of the high number of patients attending primary care and the uncertainty caused by undifferentiated symptoms that patients tend to present with (Singh, Schiff, Graber, Onakpoya, & Thompson, 2017). It is likely that current estimates of missed diagnostic opportunities across the entire health system based on medical record analysis have underestimated the problem because of the lack of data from primary care (Singh et al., 2013; Singh, Meyer, & Thomas, 2014). It is concerning that there is so little data and understanding of diagnostic error in primary care, given that primary care is where most patient contact takes place in the UK and where

failures in diagnosing patients can have serious implications for patients (DoH, 2000; Kostopoulou et al., 2008).

A recent review of primary care electronic medical records in the UK has begun to shed light on the occurrence of missed diagnostic opportunities in primary care. It estimated that missed opportunities occurred in 61.4% of consultations over a 12-month period (Cheraghi-Sohi et al., 2018; Cheraghi-Sohi et al., 2015). Moreover, the consequences of the missed opportunities were significant for over half of those that experienced an error (Cheraghi-Sohi et al., 2018). A systematic review of studies measuring diagnostic error in primary care revealed that misdiagnosis occurred for a wide range of illnesses that are difficult to diagnose because of the way they present. Diseases that presented atypically, had non-specific presentations, very low prevalence and where co-morbidity was present, were more likely to be subject to missed diagnostic opportunities (Kostopoulou et al., 2008). Primary care is a setting in which diagnostic possibilities are being negotiated rapidly, frequently and under time and resource constraints. It is, therefore, a context in which missed diagnostic opportunities are likely to occur. The wider project team made the decision to set the online resource within primary care and this thesis will focus on the evaluation of an online intervention to reduce missed diagnostic opportunities in this setting.

1.1.3 Cancer

Cancer has been identified among the diseases that present a particular risk to missed opportunities in primary care (Kostopoulou et al., 2008; Singh et

al., 2017). This is in part because the diagnosis of any one cancer in primary care is a relatively rare event; a GP may only see one case of the common cancers like lung cancer each year (Rubin et al., 2015). Furthermore, the symptoms of cancer are also symptoms found for many other more common and benign conditions (Hamilton, Peters, Round, & Sharp, 2005). Missed opportunities and delayed diagnosis have been indicated as the cause of the UK's poor survival rates for cancer, in comparison to other European countries (O'dowd et al., 2015). This is a particular issue for lung cancer, as lung cancer is the leading cause of cancer death (Bradley, Kennedy, & Neal, 2018). Lung cancer is predominantly identified at an advanced stage through emergency presentation in secondary care, which is associated with poorer outcomes for patients. It tends to present with relatively indistinct symptoms such as cough (Bradley et al., 2018; Rubin et al., 2015; Walters et al., 2013). Additionally, evidence from the wider research group, which my PhD was part of, indicated that GPs face difficulties when making diagnostic decisions where patients present with symptoms that might be lung cancer (Sheringham et al., 2016). Other cancers and diseases that can present with vague and non-specific symptoms, such as myeloma, infections and cardiovascular disease, are also commonly linked with missed diagnostic opportunities in general practice (Miller & Levy, 2015; Rubin et al., 2015; Shephard et al., 2015; Singh et al., 2017). Consequently, my thesis will primarily discuss the development of an evaluation of on an online learning tool that aims to address some of the issues related to the missed diagnostic opportunities for lung cancer in general practice. However, given that other serious conditions, can also be missed in general practice it may be possible and helpful to broaden the focus of the learning to

diagnosis and management of commonly missed serious conditions more generally, including lung cancer.

1.2 Policies to reduce missed diagnostic opportunities

Given the significant consequences of missed diagnostic opportunities on patient safety and healthcare costs, policy makers have made the reduction of diagnostic error and early diagnosis, particularly for cancer, a priority in healthcare policy (Illingworth, 2015; Yu A, 2016). In the UK, the government's mandate to The National Health Service (NHS) England for 2016-2017 included an objective to significantly reduce avoidable deaths and to improve early diagnosis for cancer patients (DoH, 2015). Furthermore, the recent *NHS Long Term Plan* commits to improving cancer survival by increasing the proportion of cancers diagnosed early from a half to three quarters by 2028 (NHS, 2019). Internationally, the World Health Organisation (WHO) has also prioritised improving safety in primary care (Sheikh, Panesar, Larizgoitia, Bates, & Donaldson, 2013). The WHO have produced a *Technical Series on Safer Primary Care*, which describes strategies and actions policy makers and health professionals can take to improve patient safety in primary care (WHO, 2017). However, the IoM's latest report, *Improving Diagnosis in Health Care*, argued that the problems of diagnostic error had been given little attention since their initial landmark report in 2000, *To Err Is Human: Building a Safer Health System* (IoM, 2000).

The IoM's latest report urged researchers, healthcare professionals and policy makers to address the issue of missed diagnostic opportunities using a

variety of strategies, one of which is to improve the teaching of clinical reasoning skills in medical education (IoM, 2015). Furthermore, they recommended that such training should be provided throughout a clinician's education, including in the earliest stages of learning as a medical student (IoM, 2015). It is important to provide training to students on the process of clinical reasoning as early as possible, as it is harder to develop these thought processes in the later stages of training (Audétat et al., 2012). Medical educators have also recognised that explicit training on how to use clinical reasoning skills in a variety of clinical contexts is currently lacking (Cleland, Abe, & Rethans, 2009; Higgs, Jones, Loftus, & Christensen, 2008; Page, Matthan, Silva, & McLaughlin, 2016). The wider research team also conducted market research and a steering group with medical educators and clinicians that indicated that there was a need for further clinical reasoning training. They indicated that medical students could be more receptive and benefit more from an online resource to improve their reasoning than professional clinicians. Providing training on reasoning earlier in the career development of clinicians could be a more effective strategy for developing reasoning skills, than attempting to retrain thought processes of more experienced clinicians (Audétat et al., 2012). Thus, the focus of this thesis will be to develop and evaluate an intervention that could be used to provide training to undergraduate medical students on how to use clinical reasoning skills in the context of primary care. **Section 1.4 and 1.5** will further explore and define my understanding of clinical reasoning and **1.6** will discuss how using a simulated online educational intervention could be well suited to providing such training.

1.3 Causes of missed diagnostic opportunities

In order to address the directives from policy makers to reduce missed opportunities in diagnosis, researchers have attempted to understand their causes. The causes fall into three categories: no-fault, system-related and cognitive or clinical reasoning errors (Graber, Franklin, & Gordon, 2005). No-fault errors include patient delay in presenting to health professionals. System-related errors include technical failures and organisational problems. Cognitive errors include errors in the thought processes of the clinicians involved in diagnosis, such as having incorrect or insufficient knowledge, failing to conduct essential diagnostic tests or ask the patient for relevant information and failing to correctly interpret information (Graber et al., 2005).

Graber and colleagues found in their assessment of medical records, that out of 100 patients who experienced diagnostic error, only seven cases experienced no-fault errors, 65% involved system-related errors and 74% experienced cognitive errors (Graber et al., 2005). Faulty data gathering, data synthesis and information processing were identified as the most common cause of cognitive error. A further study also found that most errors (78.9%) in primary care were caused by process breakdowns in the patient-practitioner encounter (Singh et al., 2013). The most common errors were related to cognitive errors concerning the gathering and interpretation of data (Singh et al., 2013). The results of these studies should be treated with a degree of caution, as the causes of missed diagnostic opportunities are hard to identify retrospectively from medical records, due to the lack of detail about the context in which the error occurred. Nevertheless, they show that missed diagnostic

opportunities are largely due to errors in clinical reasoning (see **Section 1.4** for further discussion on clinical reasoning).

Prospective studies that provide evidence from observing clinicians and students' behaviour in experimental settings have also shown that clinical reasoning errors can cause diagnostic errors. For example, Kostopoulou and colleagues found that 54% of clinicians presented with patient scenarios significantly distorted information to fit their initial diagnosis, despite receiving information that conflicted with the initial diagnosis (Kostopoulou, Russo, Keenan, Delaney, & Douiri, 2012). Sheringham and colleagues also observed that GPs failed to gather the relevant data to make an informed referral of lung cancer in 42% of online patient vignettes (Sheringham et al., 2016). Medical students are also subject to missed diagnostic opportunities due to faulty knowledge and skills, including interpretation of tests and premature closure (Braun, Zwaan, Kiesewetter, Fischer, & Schmidmaier, 2017). It is important to acknowledge that findings from vignette studies may not reflect real clinical practice but they show that errors in clinical reasoning contribute to missed diagnostic opportunities, irrespective of system and no-fault errors (Norman et al., 2016). Thus, evidence suggests there is a need to improve clinical reasoning in current and future clinicians involved in diagnosis.

1.4 Clinical reasoning skills

1.4.1 Definition

Clinical reasoning can be broadly defined as the thought processes required to identify likely diagnoses, formulate appropriate questions and reach

clinical decisions (Higgs et al., 2008). However, definitions of clinical reasoning have varied widely in the literature. There is no single definition that represents clinical reasoning across different contexts and health professions (Higgs et al., 2008; Norman, 2005). Many descriptions of reasoning fail to encompass all the specific elements of clinical reasoning, such as cognition, metacognition and knowledge, and how these interact with the context of the clinician. Additionally, the term clinical reasoning is often used interchangeably with related terms, including 'problem-solving', 'critical thinking', diagnostic reasoning', 'clinical judgement' and 'decision-making' (Simmons, 2010). Clinical reasoning may involve some of these skills some of the time, but they do not sufficiently describe or equate to clinical reasoning. For example, terms such as 'decision-making' and 'judgement' might refer more to the outcomes of clinical reasoning rather than the thought processes required to reach that outcome. Additionally, 'critical thinking' may adequately describe the process of analytical clinical reasoning but it does not describe how clinicians may use non-analytical thinking, which has been proven to be an essential skill in clinical reasoning (Norman & Eva, 2010; Norris, 1985). The lack of clarity in the definition is an issue for observing and measuring clinical reasoning skills, as it is difficult to know what should be observed and measured. In order to evaluate the effect an intervention could have on such skills, it is necessary to fully define and conceptualise what is meant by clinical reasoning skills. Therefore, I have explored models of clinical reasoning (**Section 1.4.2**) that elucidate the various components of this complex process.

1.4.2 Models of clinical reasoning

Models that describe in detail the elements understood to be involved in clinical reasoning skills and how these elements interact can provide more insight into what specific skills clinical reasoning involves, as well as help to conceptualise clinical reasoning skills.

1.4.2.1 *Higgs et al. (2008) model of clinical reasoning*

Higgs and colleagues' model of clinical reasoning (**Figure 1-2**) describes it as an evolving and cyclical process that involves applying medical knowledge, gathering necessary information from a patient and other sources, interpreting (or reinterpreting) that information and problem formulation (or reformulation) (Higgs et al., 2008). The core dimensions of reasoning they describe are: **cognition**, which refers to thinking skills involved in reasoning, such as analysing and evaluating patient information and hypotheses; **reflection**, which is the ability to think about a situation and learn from it by planning to take action; **metacognition**, which refers to the ability to select, monitor and evaluate one's own thinking; the **clinical problem**, which is the collection of symptoms the patient presents with; **knowledge**, which refers to discipline content-specific knowledge and knowledge gained from experience; the **environment,** in which the reasoning takes place and the **client's input** (Flavell, 1979; Sandars, 2009). Higgs' et al. (2008) model describes the complexity of factors involved in clinical reasoning but because of its complexity, it is hard to operationalise and consider how reasoning can be observed. It also does not describe in detail the role of the different elements. For example, it asserts that cognition is important but it

does not describe what thinking strategies clinicians may use to reach a decision, such as non-analytical thinking.

Figure 1-2 Higgs and colleagues' (2008) model of clinical reasoning
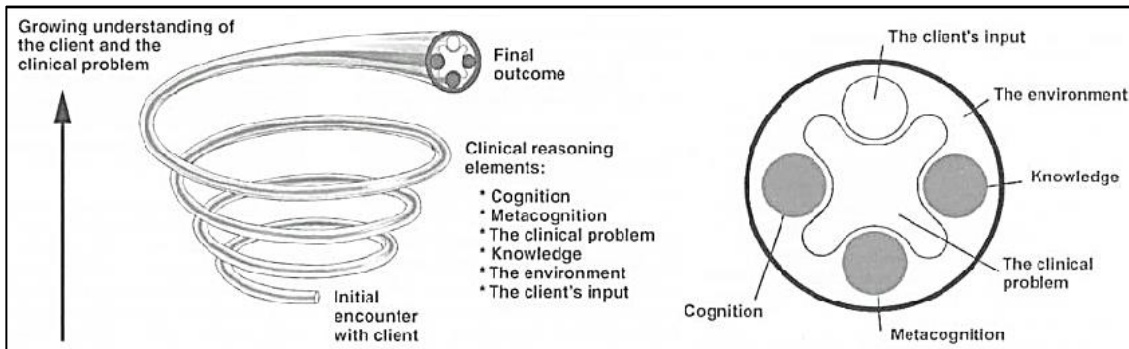
Note: image taken from Higgs et al. (2008).

1.4.2.2 Simmons' model of clinical reasoning in nursing

Simmons' (2010) conceptual analysis of the definition of clinical reasoning from a literature review of the nursing literature offered specific insights into the thinking strategies and current understanding of clinical reasoning. Simmons (2010) concluded that the definition of clinical reasoning in nursing was a "*complex cognitive process that uses formal and informal thinking strategies to gather and analyse patient information, evaluate the significance of this information and weigh alternative actions. Core essences of this concept include cognition, metacognition and discipline specific knowledge*". By informal thinking strategies, Simmons (2010) refers to the use of heuristics, pattern recognition or **non-analytical thinking**. Formal thinking strategies refer to **analytical thinking**, including the hypothetico-deductive method described by Elstein, Shulman, and Sprafka (1978), in which clinicians go through a process of generating hypotheses and gathering data to test these

hypotheses. Simmons (2010) also highlights that clinical reasoning is a **context-dependent** skill that depends on domain-specific **knowledge**, but also depends on the level of uncertainty or **risk** in a situation. Simmons (2010) also describes **cognitive flexibility**, which is the ability to use a variety of thinking strategies and adapt to different contexts and levels of risk.

Simmons' model of clinical reasoning expands on the model by Higgs et al. (2008) to offer insight into the different thinking strategies involved in clinical reasoning. However, it has not incorporated the latest models of clinical reasoning from medicine, which have gone further in describing the role of non-analytical and analytical thinking for example. Additionally, while the medical literature focuses on the diagnosis as the outcome of clinical reasoning, this may not be true for other health professions such as nursing (Norman, 2005; Simmons, 2010). As the focus of this thesis is clinical reasoning in medicine to prevent missed opportunities in diagnosis, it is important to refer to models of clinical reasoning in which the primary goal is to reach an appropriate diagnosis.

1.4.2.3 Role of mental representations in clinical reasoning models

The medical literature on clinical reasoning skills has been influenced by cognitive psychology in that it has focused on how clinicians develop mental representations, illness scripts or schemas to help them resolve clinical problems (Higgs et al., 2008). Previous research has demonstrated that clinical reasoning is more complex than simply the process of hypothesising and testing diagnoses, as described by the hypothetico-deductive model (Elstein et al., 1978; Groen & Patel, 1985). Reaching timely and accurate diagnostic

assessments are more associated with a clinician's understanding of a problem, the organisation of knowledge and the complexity of their mental representation of the problem, than hypothesis testing (Elstein et al., 1978; Norman, 2005). Additionally, experts do not always use hypothesis testing to determine a diagnosis, particularly in familiar cases. Many use pattern recognition and automatic retrieval of knowledge from their complex mental representation of problems to quickly and accurately diagnose patients (Groen & Patel, 1985). Mental representations of problems are domain specific, so good performance on one problem does not necessarily translate to good performance on another. This could explain why clinical reasoning ability varies across contexts and depending on the content of the problem (Higgs et al., 2008). The implications of this are that clinicians need to use both analytical (hypothesis testing) and non-analytical (pattern recognition) thinking strategies when engaging in clinical reasoning. The use of analytical and non-analytical thinking strategies depends largely on the clinician's content-specific knowledge and the complexity of their mental representation or understanding of the problem.

1.4.2.4 The dual-process model of clinical reasoning

The medical literature has also proposed the dual process model of clinical reasoning, which recognises a role for both analytical and non-analytical processes in clinical reasoning. The dual-process model, which originates from psychology, argues that everyone makes decisions using two systems (Kahneman, 2003; Kahneman, Frederick, Kahneman, & Frederick., 2002). System 1 (non-analytical thinking) involves quick, intuitive processing, using heuristics and pattern recognition, and system 2 (analytical reasoning) involves

reflection, flexible thinking and slow processing (Croskerry, 2009b; Kahneman, 2003; Kahneman et al., 2002). This model, as applied to clinical reasoning, proposes that when clinicians recognise symptoms they tend to use non-analytical thinking but often need to use analytical thinking in uncertain situations (**Figure 1-3**) (Croskerry, 2009a, 2009b). Non-analytical thinking is beneficial to clinicians, as it allows them to make generally accurate decisions quickly (Croskerry, 2009a, 2009b). However, an over-reliance on non-analytical thinking can lead to errors in decision-making because it is influenced by cognitive biases (Blumenthal-Barby & Krieger, 2015; Croskerry, 2009a; Eysenck & Keane, 2010).

Figure 1-3 Dual process model of clinical reasoning (Croskerry, 2009b)
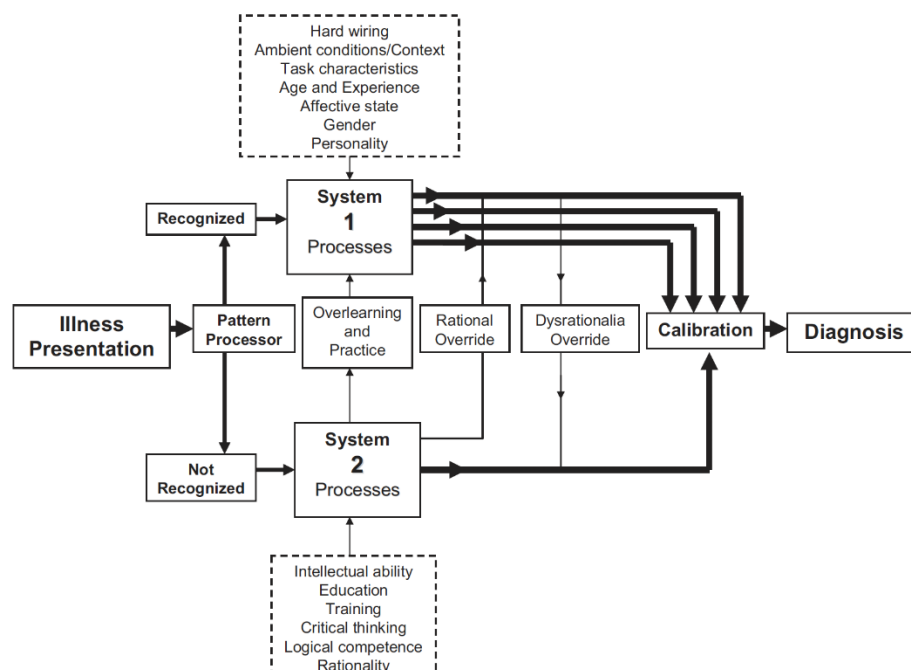
Note: image taken from Croskerry (2009b)

The dual-process model suggests that the most effective way to reduce missed diagnostic opportunities is to mitigate the effect of cognitive biases on clinical reasoning (Norman & Eva, 2010). Cognitive biases are inherent errors in thinking that deviate from rational thinking (Norman & Eva, 2010). There are many different cognitive biases that are known to influence thinking and potentially cause missed diagnostic opportunities. The following examples of biases are ones that have been identified in experimental studies as potential causes of missed diagnostic opportunities (Croskerry, 2002; Kostopoulou et al., 2012; Sheringham et al., 2016). Confirmation bias refers to the tendency to seek information to confirm a hypothesis rather than refute it; anchoring refers to the tendency to stick to an initial hypothesis despite new contradictory information and the unpacking principle refers to the tendency not to elicit all the necessary information to make an informed judgement (Norman & Eva, 2010). The dual process model suggests that analytical reasoning can help to overcome the missed diagnostic opportunities caused by cognitive biases by facilitating reflection and metacognition, which can highlight any errors made. However, caution should be taken when interpreting this model of reasoning, as one of the main criticisms of the dual process model is that it implies non-analytical reasoning may be inferior to analytical reasoning (Eva, 2005; Norman, 2005; Norman & Eva, 2010). Several studies have found that clinicians and students have greater diagnostic accuracy if they engage in non-analytical thinking (Monteiro & Norman, 2013; Norman, 2005; Norman et al., 2014; Norman et al., 2016). Clinical reasoning is a context and content-specific skill; therefore, analytical reasoning is only likely to be a more useful strategy than non-analytical reasoning if the situation is complex or uncertain and requires further

consideration of the information, which in turn will depend on factors such as the knowledge and experiences of the clinician or student.

Another criticism of the dual process model is that it implies that the clinical reasoning process is somewhat linear, which other models such as Higgs' et al. (2008) model have emphasised is not the case. It fails to show how information gathered from the patient affects the different types of thinking and changes how the perception of the illness evolves through these processes (Eva, 2005). The upward and outward spiral described by Higgs et al. (2008) may be a more appropriate representation of the evolving nature of clinical reasoning. Additionally, it is unlikely that different types of thinking are selected simply based on whether the symptoms are recognised or not. The complexity of a clinician's mental representations of similar problems will also have an effect on reasoning. Just because a clinician recognises symptoms it does not necessarily mean that they have a complex mental representation or great understanding of such problems, and that they can quickly and accurately diagnose a patient. For example, a medical student or clinician with limited experience and knowledge of particular symptoms might be able to recognise those symptoms and generate likely diagnostic hypotheses. However, because of their limited knowledge and experience they would not have a complex mental representation to draw on; consequently, they may engage in analytical thinking to ensure that they do not incorrectly interpret the information they find or not gather appropriate information to be able to rule out possible life-threatening conditions like lung cancer. Furthermore, the dual process model may not fully reflect the clinical reasoning process of medical students. For

example, medical students would likely need to rely on analytical thinking rather than non-analytical thinking much more than clinicians because of their more limited clinical experience, knowledge and less complex mental representations of problems (Croskerry, 2009b; Eva, 2005). Equally, the model does not emphasise the importance of the wider context in which the diagnosis is being made, and in particular how the level of risk and uncertainty in a problem is likely to affect the thought strategies that are used.

1.4.2.5 Universal model of clinical reasoning

Eva (2005) has proposed a combined model of clinical reasoning that more explicitly acknowledges the interaction between types of thinking in clinical reasoning (see **Figure 1-4**). It focuses less on identifying what specific thought processes could lead to errors and proposes a model of how clinicians should use their clinical reasoning skills. It suggests a bi-directional additive model, in which the clinician forms a mental representation of a patient's problem. This leads to hypothesis testing, which then changes their mental representation and the perception of a patient's problems. Analytical and non-analytical reasoning occur throughout this process, with non-analytical reasoning more likely to play a larger role in developing the initial hypotheses and analytical reasoning a larger role in testing these hypotheses. They also acknowledge there is an interaction between these two types of thinking so that one does influence the other.

Figure 1-4 Combined model of clinical reasoning (Eva, 2005)
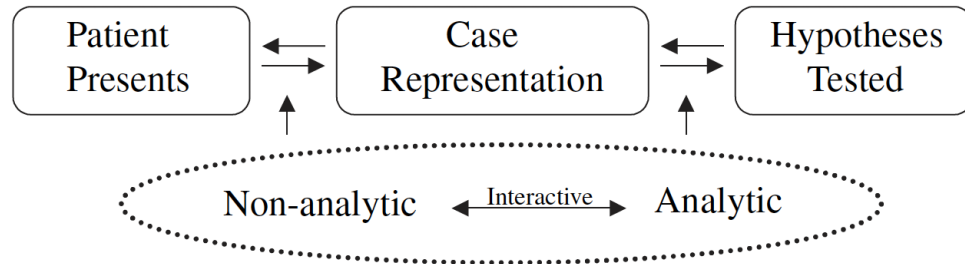
Note: image taken from Eva (2005)

1.4.2.6 Summary of clinical reasoning models

Each of these models provides insight into how clinicians and medical students reason in clinical situations. However, none of these models provide sufficient detail to elucidate all the different processes, behaviours and outcomes involved in clinical reasoning. Not all of the models have incorporated the dual process model and the influences of context, knowledge and mental representations on reasoning. A theoretical model of clinical reasoning that incorporates all of these factors is needed to fully encapsulate what clinical reasoning skills are and how they might be observed. This will help to determine what strategies could be effective at improving reasoning and how these might be measured.

Given that previous models have indicated that knowledge and previous clinical experiences influence reasoning, the theoretical model may differ between medical students and professionals. My model will focus on the clinical reasoning skills of medical students, as opposed to healthcare professionals, because they have limited knowledge and clinical experience. Furthermore,

research has highlighted the importance of formulating clinical reasoning skills early in medical education, as it can be difficult to retrain thought processes once they have become ingrained (Audétat et al., 2012). Developing 'good' thinking skills from an early stage in education could help future doctors to form 'good' clinical reasoning skills that they can apply in clinical practice and potentially reduce future missed diagnostic opportunities. In **Section 1.5,** I describe my theoretical model of clinical reasoning that will incorporate the important elements of clinical reasoning identified above as they apply to medical students.

1.5 Theoretical model of medical students' clinical reasoning

I developed a theoretical model of clinical reasoning skills to bring together the learning from models of clinical reasoning described in **Section 1.4.2** (see **Figure 1-5**). It summarises my understanding of medical students' clinical reasoning.

Figure 1-5 Theoretical model of medical students' clinical reasoning
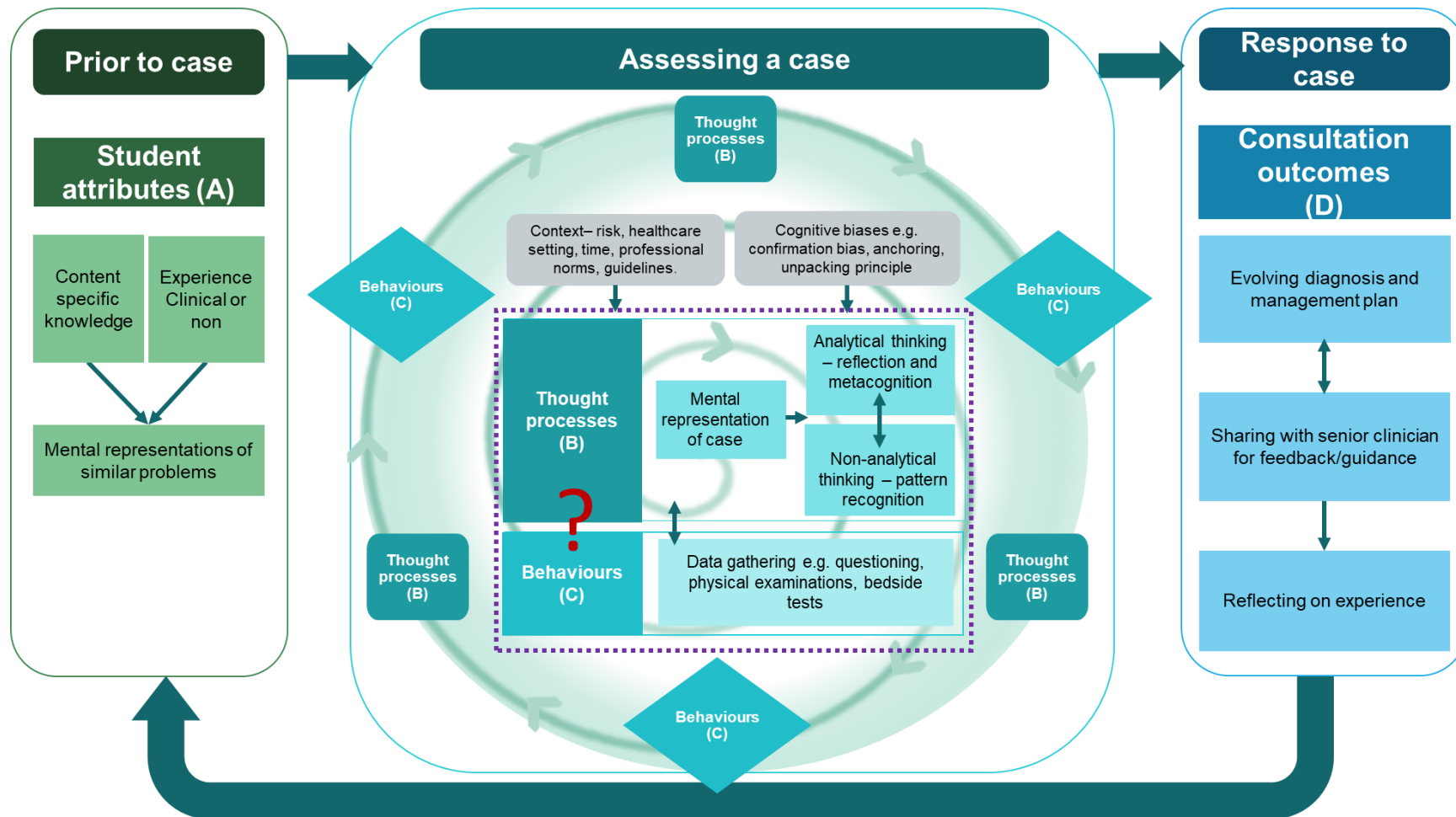
In my model (**Figure 1-5**), I have distinguished between four components of clinical reasoning that occur before a student sees a patient, during their assessment of their current patient and as a result of assessing their current patient:

- **A** – Student attributes (in green in **Figure 1-5**): the attributes that a student brings with them to their current patient case, such as their content-specific knowledge of the problem and their prior experiences, which contribute to their mental representations of illnesses of similar problems;

- **B** – Thought processes (in dark turquoise): a student's mental representation of the current patient case and their use of analytical thinking, which may include the testing of diagnostic hypotheses, metacognition and reflection; and their use of non-analytical thinking, which may include hypothesis generation and pattern recognition;

- **C** – Behaviours (in light turquoise): when a student gathers data, which includes questioning, physical examination and undertaking bedside tests in order to test their diagnostic hypotheses;

- **D** – Consultation outcomes (in blue): a student's evolving diagnosis and management plan that they then share with a senior clinician to receive feedback on their clinical reasoning, decisions and management. They may reflect on this experience, which feeds back into their student attributes (mental representations of similar problems) for future cases.

- Red question mark – Signifies that this thesis will focus on the link between thought processes and behaviours in reasoning.

The cognitive biases of confirmation bias, anchoring and unpacking principle that have been found to affect clinical reasoning have been included as factors that influence the thought processes and subsequent behaviours of the students. Contextual factors in which the reasoning takes place, such as the level of risk, time pressure, professional norms and clinical guidelines, are also included as factors that affect student thought processes and behaviours. It should be noted that this cycle can be repeated multiple times for any given patient, as diagnoses often evolve over time depending on symptoms, test results and response to treatment.

My understanding of the process of clinical reasoning for medical students is that clinical reasoning begins before a student even sees a patient. A student's previous experiences, both clinical and non-clinical, and their content-specific knowledge of this type of problem, have helped them already to form an understanding of similar problems to what they are presented with (otherwise known as their mental representations or schemas of similar problems). When a student begins to assess a new patient case, they create a mental representation of that specific case based on their mental representations of similar problems and their use of analytical or non-analytical thinking. Both thinking strategies could be used in the same patient case. Non-analytical reasoning might be used by the student at the beginning of the case to generate hypotheses and recognise patterns in the information presented.

Analytical reasoning may be used later in the case to test their hypotheses and they may engage in reflection and metacognition to help them refine their diagnoses. Use of the different thinking strategies also depends on the complexity of their mental representation of the case. A student, who has a lot of knowledge about the presenting problem and has experienced many patients with these symptoms may already have a complex mental representation of similar problems; consequently, they may have a more complex representation of the current case and may be more likely to use non-analytical thinking to reach a diagnosis quickly. Students might be more likely to use analytical reasoning than professionals because they have relatively limited clinical experience and, therefore, will be unlikely to already have a complex mental representation of similar problems to draw on. Certain components of clinical reasoning cannot be directly observed, such as a student's thought processes and mental representations. Therefore, the description of this component may not be complete. However, insight into a student's thought processes during a case could be reflected in a student's behaviours or actions during a case, which can be observed.

Observing what information students gather from a patient and why they are gathering that information could help to elucidate the thought processes behind their actions. Gathering data from a patient requires students to ask questions, physically examine the patient and order appropriate tests. There is an interactive and bi-directional relationship between the thought processes (and use of analytical and non-analytical thinking) and behaviours involved in clinical reasoning, with one influencing the other, until a satisfactory diagnosis

is reached. For example, if a diagnosis becomes apparent based on the information gathered then non-analytical thinking might be used and result in the end of data gathering and hypothesis testing, as the information gathered is sufficient to make a suitable and quick diagnosis. In cases where there is uncertainty or unexpected information is gathered, this may influence the student to use analytical thinking and engage in further hypothesis testing and investigation to reach a diagnosis.

Cognitive biases could influence both thinking strategies, potentially causing errors in a student's thinking, some of which may manifest in observable behaviours, such as only seeking information to confirm their hypotheses (confirmation bias); sticking to their initial diagnosis (anchoring); or not seeking the necessary information to make an informed diagnosis (unpacking principle). Clinical reasoning thought processes and behaviours are also dependent on the context in which they are set. Factors, such as the level of risk the clinical situation presents; the type of healthcare setting (e.g. primary or secondary care); the time restraints placed on the students, and practice guidelines and professional norms for diagnosing and managing patients with these symptoms, can all affect clinical reasoning. For example, in clinical scenarios where there is high risk, analytical thinking strategies may be used over non-analytical to ensure that the judgement is appropriate, as the consequences of an error could be significant.

The consultation outcomes of these thought processes and behaviours that occur in clinical reasoning are the evolving diagnosis and management plan

for the patient. While students may consider several diagnoses during their thought processes to guide their data gathering, their diagnosis and management plan refers to the decision they reach regarding diagnosis at the end of the consultation. Students will present their clinical judgement and reasoning to a senior clinician for feedback and guidance on how to diagnose and manage the patient. This feedback provides students with an opportunity to reflect and learn from their experiences, which adds to their knowledge, experience and mental representations of similar problems that they bring with them to their next case.

Overall, my theoretical model seeks to clarify the different elements of students' reasoning processes and how they could be measured. Further empirical and theoretical evidence is needed to inform how the reasoning skills of students can be improved to address the clinical and educational need to improve clinical reasoning skills training. This will be described in my logic model in **Chapter 4**. My theoretical model informs the logic model by indicating what the targets of an intervention to improve reasoning could be, such as influencing the mental representations of students, which in turn could improve behaviours and consultation outcomes. It also shows that the thought processes of analytical and non-analytical reasoning are both important and need to be developed to improve reasoning. Additionally, this model suggests how underlying and unobservable thought processes can be observed via behaviour and the consultation outcomes. It is important to make these assumptions clear to show how clinical reasoning skills could be measured in evaluations.

1.6 Educational strategies to improve clinical reasoning skills

The models of clinical reasoning described in **Section 1.4.2** not only informed my understanding of the theoretical model of medical students' clinical reasoning, they also suggested some ways of improving students' reasoning, such as by facilitating analytical reasoning. Evidence has suggested that there is currently a gap in undergraduate medical education to deliver explicit teaching on clinical reasoning skills, which could have an impact on their clinical reasoning skills as future doctors (Cleland et al., 2009; Higgs et al., 2008; Page et al., 2016). However, there are significant barriers to delivering explicit training on reasoning in undergraduate medical education, such as the difficulties of getting sufficient patient contact in a variety of contexts and the increased burden on faculty's time and resources to create and deliver high quality and consistent clinical reasoning teaching. Simulated patient cases delivered online could surmount these barriers and, thus, enable schools to teach clinical reasoning skills explicitly. Indeed, theories of learning and cognition also support the use of simulations for learning complex skills, such as reasoning. The recent Topol review, on preparing the healthcare workforce to deliver the digital future, emphasises that a wide variety of learning methodologies, including technology enhanced learning, should be used in conjunction with face-to-face experiences to teach clinical skills to students. Delivering teaching using a variety of learning methodologies, including technology-enhanced learning, could help to fill gaps in the provision of teaching, reach a wider number of students, create more personalised learning experiences and have the possibility of tracking life-long learning via technology (Topol, 2018).

Nevertheless, it is important to acknowledge that use of technology-enhanced learning in medical education has its own limitations, such as the lack of engagement from users, lack of fidelity with real patient consultations and technical problems. Further understanding is also needed on how these technologies can best fit into curricula, complement current training and how effective they are.

1.6.1 Current teaching of clinical reasoning skills in undergraduate education

A survey of almost half of the undergraduate medical schools in the UK, found that the most common way of teaching clinical reasoning was through small group tutorial or problem-based learning, followed by clinical or communication skills sessions and primary care placements (Page et al., 2016). Much of the training for clinical reasoning relies on exposure to real patients (Schmidt & Mamede, 2015). Learning clinical reasoning skills through interaction with real patients is seen as the 'gold-standard' of teaching reasoning, as it gives students the opportunity to practise their reasoning in real-life and get feedback from a senior clinician on their skills. However, there are limitations with this approach, as supervision and feedback on performance is often variable and there is limited time for students to reflect actively on their skills and improve (Schmidt & Mamede, 2015). Moreover, organising and delivering teaching with real patients is time and resource intensive, which may hinder the development of more explicit teaching on clinical reasoning.

Students also have limited exposure to a large number and wide variety of patients in different contexts to practise their clinical reasoning skills (Harding, Rosenthal, Al-Seaidy, Gray, & McKinley, 2015; Schmidt & Mamede, 2015). Exposing students to cases in primary care may be a particular issue; a recent cross-sectional study of UK medical schools revealed that students received only 13% of their clinical training in general practice (Harding et al., 2015). Therefore, medical students may have limited opportunities to practise applying their clinical reasoning skills in primary care contexts. It is important for students to have a good understanding of clinical reasoning in primary care contexts and the challenges of reasoning in such complex settings. Primary care is a context where missed diagnostic opportunities are highly likely to occur and there are growing demands to deliver care in general practice (Harding et al., 2015; Kostopoulou et al., 2008).

To recognise the limitations of teaching clinical reasoning skills through real patient contact, medical schools have developed ways of providing students with more opportunities to practise their clinical skills (Schmidt & Mamede, 2015). This additional training is most commonly presented via prefabricated clinical cases. These can be paper-based, virtual or with actors representing patients and can be used in small group discussions or individually (Schmidt & Mamede, 2015). Some use simulation techniques that are designed to represent a situation that a learner would face in their real working environment and require the student to actively acquire data and respond as they would in real-life, whereas some may be more static and provide all patient information upfront (Issenberg, Mcgaghie, Petrusa, Lee Gordon, & Scalese,

2005; Maran & Glavin, 2003). A common example of teaching using patient cases in the UK and worldwide is problem-based learning (PBL). PBL involves students solving authentic patient problems in small groups with the guidance of a facilitator (Barrows, 1994; Norman & Schmidt, 1992). PBL has been found to be effective at improving knowledge and clinical skills but is still time and resource intensive for medical schools to deliver (Norman & Schmidt, 1992; Schmidt & Mamede, 2015). Furthermore, PBL does not always simulate a clinical scenario as there is not always interaction with a simulated patient or requirement to gather information, so they may not be suitable for teaching all aspects of clinical reasoning skills like data gathering.

1.6.2 Patient simulations

Simulated patients can be used for both teaching and assessing clinical reasoning skills in undergraduate medical education (Higgs et al., 2008). They usually involve actors who have been coached to simulate a real patient and are used to help provide students with patient contact as early as possible (Cleland et al., 2009). The main benefit to simulations is that students can practise skills, such as data gathering and making diagnostic decisions, without endangering patients (Issenberg et al., 2005). They enable students to be observed by experts, so their actions can be assessed and immediate feedback can be provided (Croskerry, 2002; Issenberg et al., 2005). Furthermore, simulations can be specifically designed to provide realistic practice for clinical reasoning in specific contexts that they may have little access to or experience of, such as complex cases in general practice (Issenberg et al., 2005). They can also facilitate the development of analytical clinical reasoning skills by

allowing consultations to be paused and giving students opportunity to reflect and receive feedback in real-time.

Simulations have been found to be more effective than didactic teaching for learning clinical skills and have been proven to be an effective method of improving skills in other sectors, like the military and the aviation industry (Cleland et al., 2009; Issenberg et al., 2005). There are, however, significant disadvantages to using simulated patients, such as the substantial costs involved in training and hiring the 'patients'; the significant time and resources it takes to facilitate the learning; the difficulty of providing this training to learners who are dispersed across geographical areas and the limited access to actor patients (Cleland et al., 2009).

1.6.3 Delivering patient simulations online

Patient simulations that are delivered online could address the significant practical and resource issues of delivering further training on explicit clinical reasoning using patient simulations. Online patient simulations are defined as *"a specific type of computer-based program that simulates real-life clinical scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions"* (Association of American Medical Colleges, [AAMC], 2007). They have similar benefits to simulated patients but are lower in cost in the long-term, as they can be reused and potentially require less time from faculty in delivering face-to-face teaching. They can also be distributed widely, completed remotely,

tailored to the learner and frequently updated (Issenberg et al., 2005; Vaona et al., 2015).

Currently, medical students have little access to online patient simulations and they are not widely used to teach clinical reasoning (Cook & Triola, 2009). However, the growing popularity of massive open online courses (MOOCs) and other online learning platforms suggests that there is an appetite for the development of online technologies to deliver training to medical students (Harder, 2013; MedBiquitous, 2017; MOOC List, 2016). Indeed, education technology companies such as MedBiquitous are now offering services to develop virtual patients to teach clinical skills (MedBiquitous, 2017). Furthermore, initiatives such as eViP (electronic Virtual Patients), funded by the European Commission, have developed a bank of 320 freely available virtual patients for medical schools to use to help teach a variety of clinical skills (eViP, 2019). While many of the cases currently available do not necessarily focus on teaching clinical reasoning, lack interactivity and lack elements of simulation (such as being able to gather data), they begin to show educators the potential of using such technology to teach clinical reasoning skills. It is important, therefore, to understand the empirical and theoretical evidence to support the use of such tools in education to ensure that they are designed and delivered appropriately. This is where my thesis can contribute to the field.

1.6.4 Empirical evidence to support the use of online patient simulations

There is evidence to support the use of simulations in medical education if they are used under the right conditions and contain appropriate features, such as inclusion of feedback, repeated practice and integration into the curriculum (Issenberg et al., 2005; McGaghie, Issenberg, Petrusa, & Scalese, 2006; McGaghie, Issenberg, Petrusa, & Scalese, 2010). There is also some empirical evidence to support the use of delivering simulations online. Two systematic reviews have found promising results that online patient simulations may improve the general clinical skills of health professionals and students, compared to no formal instruction (Consorti, Mancuso, Nocioni, & Piccolo, 2012; Cook, Erwin, & Triola, 2010). However, previous reviews found only a small proportion of studies on the effectiveness of simulations or online simulations were of high enough quality to make meaningful conclusions (Issenberg et al., 2005; McGaghie et al., 2006; McGaghie et al., 2010). Therefore, the conclusions from these reviews should be treated with caution.

No previous systematic reviews of the literature have looked specifically at the effects of online patient simulations on clinical reasoning skills for undergraduate medical students. It is important to look at the effectiveness of such technologies on medical students specifically, given that it is in undergraduate medical education where a need has been identified to improve explicit teaching of reasoning, potentially through simulation. Moreover, students and professionals differ in their level of knowledge and clinical experience, which may mean that simulations may have different effects on

their clinical reasoning skills. It is also important to understand what effect online patient simulations might have on clinical reasoning skills specifically, since it is these skills that have been found to be the main cause of missed diagnostic opportunities. Currently available reviews of the literature may also already be out of date given the fast-paced nature and growing trend in online learning. Thus, a further systematic review is needed to understand the current empirical evidence for using online patient simulations to improve undergraduate medical students' clinical reasoning skills. Given the paucity of empirical evidence that reviews have currently found, it is likely that the development and evaluation of an online patient simulation tool to improve clinical reasoning would fill the gap in the field.

1.6.5 Theoretical evidence to support the use of online patient simulations

There is considerable theoretical evidence from the fields of education and psychology that supports the use of simulations for teaching complex, context-dependent skills such as clinical reasoning, which will be discussed in this section (Cook & Triola, 2009). The implications from theories of education and psychology on simulations have substantial overlap. For example, both theories recommend improving and learning skills through simulation because simulation can facilitate reflection and feedback. More commonly, the cognitive theories have been used to describe strategies to improve reasoning in clinicians; the learning theories have been used in undergraduate medical teaching to describe ways of teaching and learning clinical reasoning skills. The following sections will discuss in more detail how these theories overlap and

support the use of simulation to improve the teaching of clinical reasoning skills of students.

1.6.5.1 Cognitive and clinical reasoning theories

Theories of cognition and clinical reasoning models described in **Section 1.4.2**, have indicated that it is the organisation of information and complexity of the mental representations of illnesses that determine expertise in clinical reasoning. As such, being exposed to a large number of different clinical cases could help to improve reasoning, as it would give students the chance to use their skills and restructure their current mental representations (Cook & Triola, 2009; Eva, 2005). This could help to improve both non-analytical and analytical thinking. Exposure to a variety of cases could help the development of pattern recognition skills and facilitate non-analytical thinking. It may also help students to identify particular types of cases or symptoms that require closer attention and more analytical thinking (Pennaforte, Moussa, Loye, Charlin, & Audetat, 2016). Delivering patient simulations online offers a practical way for medical schools to provide a large number and range of clinical cases that can be easily adapted and updated according to the needs of students

Theories of clinical reasoning also suggest that facilitating metacognition, reflection, and providing feedback can improve reasoning, as it can help students to develop their analytical reasoning skills (Audétat, Laurin, Dory, Charlin, & Nendaz, 2017; Lambe, O'Reilly, Kelly, & Curristan, 2016; Mamede et al., 2010; Norman et al., 2016; Sandars, 2009). Metacognition, reflection and feedback may help students to detect errors in their thinking skills

or knowledge during and after a task, which they can take action on and plan to improve in their next case (Mamede, Schmidt, & Penaforte, 2008). Online patient simulations could facilitate reflection and metacognition in students, if they require students to plan how they will do the task, evaluate their thinking strategies during and after the task, make decisions in uncertain circumstances, attend to feedback and explicitly reflect on how they performed in the task (Azevedo, 2005; Flavell, 1979; Schraw, 1998). Furthermore, they could also be designed to incorporate prompts that explicitly require students to engage in metacognition and reflection to ensure that these thought processes are triggered. Immediate feedback can also be provided based on their actions in the simulation, which can help to restructure their mental representations of illnesses.

1.6.5.2 Theories of learning

1.6.5.2.1 Constructivism

Medical education generally follows the principles of learning outlined by constructivist learning theory. Constructivist learning theory argues that learning is a process of constructing meaning. Learning is seen as an active process, in which new knowledge is constructed based on our understanding of experiences, in reference to our previous experiences, culture and society. It argues that learning 'by doing' leads to the restructuring and reconfiguration of current knowledge (Merriam, Caffarella, & Baumgartner, 2012; Taylor & Hamdy, 2013; Vygotsky, 1997). It, therefore, supports the use of learning clinical reasoning skills by providing students with more experiences; through

interaction with online patient simulations, students can be provided with a multitude of experiences to learn from.

1.6.5.2.2 Experiential learning theory

Sub-theories of constructivism can offer further insights into how learning by experience through online patient simulations could improve learning of clinical reasoning skills. The experiential learning theory argues that learning is the result of an individual's experiences, which they internalise and give personal meaning to (Kolb, 1984; Yardley, Teunissen, & Dornan, 2012). Kolb's model of experiential learning describes the cyclical process of learning (See **Figure 1-6)** (Kolb, 1984). In Kolb's model, the learner has an experience; they reflect on that experience; they form abstract concepts and generalisations to understand their experience; and finally, they test their understanding of the experience in new situations (Kolb, 1984). Thus, reflection on experiences and the ability to test hypotheses is key to learning and developing expertise in skills. Patient simulations allow students to: learn through experience; reflect on their actions; test their ideas and see the consequences of those actions in a safe environment.

Figure 1-6 Kolb's learning cycle
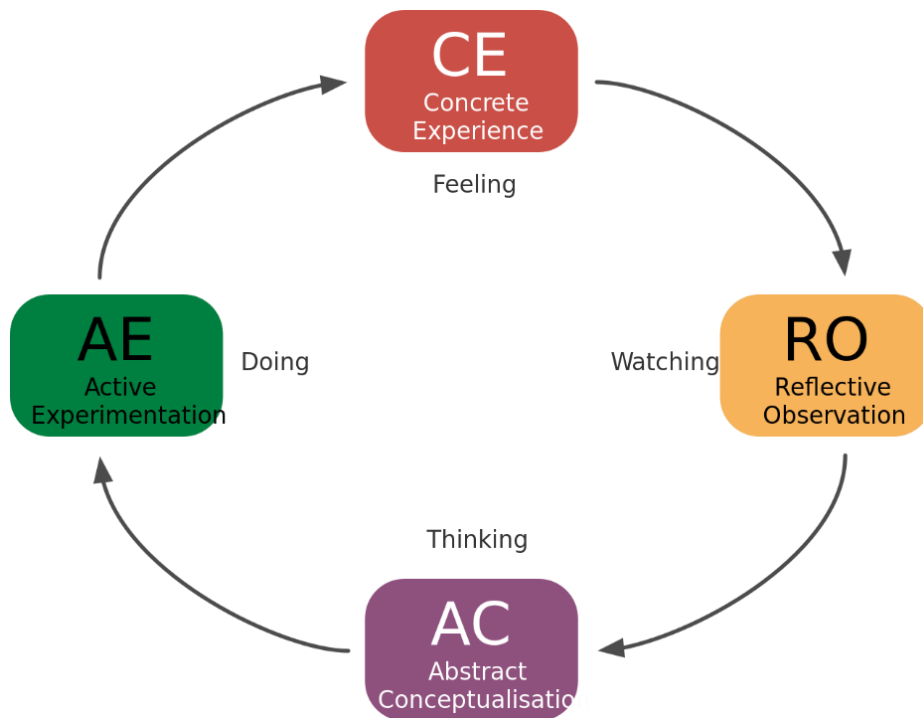
Two types of reflection can occur in experiential learning. Reflection-in-action, which is where the experience is compared to prior knowledge and experiences; differences between the two are reflected on during the task (Schön, 1983). Reflection-on-action refers to reflecting on the thought processes used to make a decision and whether they were appropriate after a task (Schön, 1983). Schön argues both types of reflection can help clinicians to be reflective practitioners and that being reflective is an important skill in clinical reasoning (Higgs et al., 2008; Schön, 1983). Online patient simulations may offer a unique opportunity to facilitate both types of reflection. Not only can reflection-on-action be facilitated online by prompting students to debrief on what they have learned, but reflection-in-action can be supported by use of prompts and pausing the simulation to allow the student to reflect during the

64

task (Pennaforte et al., 2016). It may be inappropriate or unfeasible to pause a consultation and ask the student to reflect continuously in clinical consultations with real patients.

1.6.5.2.3 Reflective practice model

The reflective practice model, which is also based on constructivist theories of learning, describes how reflection and action in learning improve performance in medicine (Mamede & Schmidt, 2004; Taylor & Hamdy, 2013). Reflection is understood to be essential in acquiring new knowledge, as reflection helps learners to construct meaning, understand their experiences and restructure their mental constructs (Mamede, Schmidt, & Rikers, 2007; Taylor & Hamdy, 2013). It is based largely on the work of Dewey, who outlined the five-stage process of reflective thought:

1   A state of uncertainty due to a complex problem

2   Understanding of the problem

3   Using inductive reasoning to suggest a solution for the problem

4   Using deductive reasoning to think of implications of their solution

5   Testing hypotheses either in real life or imaginatively (Dewey, 1997; Mamede & Schmidt, 2004).

It is also based on the concept of deliberate practice as the process by which learners acquire expertise (Ericsson, Krampe, & Tesch-Römer, 1993). Deliberate practice requires repetitive practice using critical reflection on performance. It focuses on identifying and addressing weaknesses of that performance, with the goal of improving future performance (Ericsson et al.,

1993). Mamede and Schmidt (2004) have provided empirical evidence of five different reasoning processes and behaviours of reflective practice, which were present in expert clinicians. These were: deliberate induction, deliberate deduction, testing hypotheses against the problem at hand, an attitude of openness towards reflection and meta-reasoning (Mamede & Schmidt, 2004). Online patient simulations offer an opportunity for students to practise being reflective by: creating a state of uncertainty via a complex clinical case; using specific prompts to trigger inductive and deductive reasoning; allowing students to test hypotheses by asking questions; ordering tests and doing examinations without practical constraints; identifying weaknesses and repeating these skills in further cases.

1.6.5.2.4 Scaffolding

Scaffolding is an instructional strategy supported by constructivist approaches to learning and is similar to guided reflection (Taylor & Hamdy, 2013; Vygotsky, 1997). A teacher or peer (expert) can provide scaffolding in a structured way to help guide a learner (novice) through a task. Scaffolding is essential for the progression of learning from novice to expert. It provides support to the learner on how concepts fit together in the context of a situation (Sharma & Hannafin, 2007). An expert's role is to help the student identify strategies to accomplish a task. Gradually scaffolding can be removed, as the learner is able to complete the task alone (Sharma & Hannafin, 2007; Taylor & Hamdy, 2013). Scaffolding can take different forms, for adult learners it is possible to use 'Socratic scaffolding' by prompting the learner to question and reflect on their task performance (Beyer, 1997; Sharma & Hannafin, 2007).

Scaffolding can be provided in simulations via prompts to reflect on clinical reasoning skills and by providing rich feedback (Sharma & Hannafin, 2007). Online patient simulations can provide scaffolding if it restricts user actions and the student's attention can be directed to the important aspects of a task (Sharma & Hannafin, 2007).

1.6.5.2.5 Self-regulated learning theory

Theories of learning that focus on how adults learn as opposed to children, such as self-regulated learning theory, offer insights into how simulations can be used to provide undergraduate students with engaging learning. Self-regulated learning theory, which is based on constructivism, sees learners as active agents in their learning and constructing meaning from their experiences (Kuiper & Pesut, 2004). However, it also draws on social cognitive theories of learning that emphasise the importance of motivational factors in learning, including affect regulation and self-efficacy (Bandura, 2001; Zimmerman, Boekarts, Pintrich, & Zeidner, 2000). Self-regulation requires metacognition and personal agency to regulate personal influences, such as self-efficacy or the belief that one can achieve a task, motivation, emotional processes and behavioural and environmental influences (Zimmerman, 1995). This theory highlights that any strategy to improve the learning of clinical reasoning should ensure that students are engaged in learning. Online patient simulations can improve students' engagement, motivation to learn and self-efficacy if they have: increasing levels of difficulty in the patient cases, so that students begin with easy cases and build their self-efficacy; or some

educational or monetary reward for performance or competition between other students by ranking their performances to improve motivation.

1.7 Summary

In this introductory chapter, I presented the clinical and public health importance to improve clinical reasoning skills to reduce missed diagnostic opportunities, particularly for the early detection and diagnosis of cancer in primary care. A theoretical model was derived from the literature to conceptualise students' clinical reasoning. I explored how theories of learning and clinical reasoning may inform the design of online patient simulations to deliver training on clinical reasoning skills. However, I found significant gaps in empirical evidence to demonstrate that this approach would be effective for teaching medical students. It is likely that further empirical evidence on the effectiveness of such tools, for medical students' clinical reasoning skills, is available and could aid the understanding of whether they are effective and what the required features are for such tools. Therefore, an up to date systematic review of the literature is needed to inform the logic model of how online patient simulations can improve students' reasoning skills. An evaluation of a theoretically informed online patient simulation tool that seeks to assess the feasibility and effectiveness of such tools would significantly contribute to the field.

2   Aims and objectives of this thesis

2.1   Aims

The aims of this thesis are to:

1. Evaluate a training tool based on evidence and theory to improve the clinical reasoning skills of medical students;

2. Contribute to the current literature by developing a logic model of how online patient simulations can improve clinical reasoning skills and by developing novel ways of identifying and assessing clinical reasoning skills.

2.2   Research questions

My research aims to address the following research questions:

1. What is currently known about the effectiveness of online patient simulations on the clinical reasoning skills of medical students?

2. What does the theoretical and empirical literature indicate about the required features of online patient simulations to promote medical students' clinical reasoning skills?

3. Is it feasible to use a RCT design to test the effectiveness of an online patient simulation?

   a) Are online patient simulations acceptable to medical students and perceived to improve clinical reasoning skills?

   b) Are the measures used to assess clinical reasoning reliable, sensitive and valid?

c) What are the potential impacts of an online patient simulation on clinical reasoning skills?

4. How do students reason when using an online patient simulation?

5. How do other factors, such as the design of the online patient simulation and students' confidence in their skills, affect reasoning?

6. How should an online patient simulation tool be designed to improve the clinical reasoning skills of medical students?

7. How should online patient simulation tools be evaluated in medical education?

2.3 Research objectives

My research objectives to answer my research questions were to:

1. Conduct a systematic review (**Chapter 3**) of the literature to explore the effectiveness of online patient simulation training tools on the clinical reasoning skills of undergraduate medical students (Research Question [RQ] 1);

2. Develop a logic model (**Chapter 4**) to show how an online patient simulation could improve students' reasoning skills, based on findings from my review, theory and feedback from users and experts (RQ 2);

3. Undertake a feasibility RCT (**Chapters 5 and 6**) to establish the feasibility and acceptability of implementing and evaluating an online patient simulation training tool in medical schools in the UK. I will also develop and test the validity and reliability of different measures of clinical reasoning skills and explore the potential impact of online patient simulations on clinical reasoning skills (RQ 3);

4. Use a qualitative Think Aloud and semi-structured interview study (**Chapters 7 and 8**) to explore how students reason when using an online patient simulation and the factors that affect reasoning (RQ 4 & 5);

5. Synthesise my findings in my discussion (**Chapter 9**) to share the learning from my PhD on how to design and evaluate an online patient simulation (RQ 6 & 7).

3 Systematic review and meta-analysis:

Current evidence of the intervention features and effectiveness of online patient simulations aimed at improving medical students' clinical reasoning skills

3.1 Background

In **Chapter 1** I discussed that there is a need to provide more explicit training on applying clinical reasoning skills in undergraduate medical education, and this training could reduce the impact of future missed diagnostic opportunities (Cleland et al., 2009; Higgs et al., 2008; IoM, 2015; Page et al., 2016). Online patient simulations have been suggested as a novel approach to teaching clinical reasoning skills explicitly (Bradley, 2006; Cook & Triola, 2009). They allow students to practise clinical reasoning with realistic 'patients' in a safe environment (AAMC, 2007; Cook & Triola, 2009; Issenberg et al., 2005; Mann & Roland, 2017). They may also be particularly suited to providing training on clinical reasoning skills that require deliberate practice with a wide variety and large number of clinical cases, which many students may not get access to with current standard approaches (Cook & Triola, 2009). Furthermore, there are practical benefits to using online based patient simulations to teach reasoning over traditional methods, as they can be more cost effective in the long-term, more easily updated and delivered to students who may be spread across large geographical areas. Indeed, transferring some types of clinical teaching to be delivered online may be the only way to satisfy the increasing demand for

medical education and to meet the requirements of medical bodies (Greenhalgh, 2001).

Despite perceived advantages of online simulation for this purpose, its effectiveness and optimal method of delivery remains uncertain. Previous systematic reviews and meta-analyses have indicated that online learning in general tends to be more effective than no intervention for teaching health professionals and students clinical skills (Consorti et al., 2012; Cook et al., 2010; Cook et al., 2011). However, because previous research has measured a wide range of non-specific learning outcomes, only some of which align with definitions of clinical reasoning, the conclusions of previous reviews may not be relevant. Additionally, the heterogeneity of outcome measures used to assess clinical reasoning and the poor reliability and validity of these measures, has made results from these evaluations difficult to interpret (Cook et al., 2010; Cook et al., 2011). Furthermore, both students and professionals from a wide range of health disciplines were included in previous reviews, even though the clinical skills required may vary depending on discipline. The inclusion of students and professionals in these reviews is also problematic because students (novices) may have poorer clinical reasoning skills compared to professionals (experts), as they have less knowledge and clinical experience, which may influence how effective such tools are (Ericsson, 2008; National Research Council, 2000). Importantly, previous reviews have also noted the large degree of heterogeneity in online interventions in medical education, only some of which could be defined as simulations (Consorti et al., 2012; Cook et al., 2010). The field is also likely to have moved on considerably since these

reviews, as digital innovations in education evolve quickly. It, therefore, remains unclear what specific effect online patient simulations have on undergraduate medical students' clinical reasoning skills (Consorti et al., 2012; Cook et al., 2010; Cook et al., 2008). A review that focuses on synthesising the specific effects of online patient simulations on medical students' clinical reasoning skills would help to identify whether these interventions are effective in this context, to update the evidence of previous reviews and provide evidence on how they are developed and delivered.

Many unanswered questions remain from the current literature regarding what features of online patient simulations are required for effectiveness. Previous research has indicated several features of online patient simulations that are theoretically likely to be effective in improving clinical reasoning skills (see **Section 1.6**). These include: providing scaffolding via directing students' attention to important aspects of reasoning; having increasing levels of difficulty; offering rewards or encouraging competition through anonymous peer comparison; providing repeated practice through a range of cases; encouraging collaborative learning through group work; providing feedback and prompting post-case and during-case reflection (Posel, Fleiszer, & Shore, 2009; Posel, McGee, & Fleiszer, 2015). It is unclear to what extent these features are used, how they are incorporated into online patient simulations and whether they are effective at improving clinical reasoning skills. A review that describes the intervention features of current online patient simulations will further our understanding of why online patient simulations might be effective and how to implement important features that will aid the learning of clinical reasoning skills.

3.2 Aims and objectives of the systematic review

In light of the above, I undertook a systematic review to address my first research question, which was to explore what is currently known about the effectiveness of online patient simulations on the clinical reasoning skills of medical students. My study objectives were to:

a) Identify the intervention features of online patient simulations designed to improve clinical reasoning skills;

b) Identify the outcome measures used to assess clinical reasoning skills across different clinical contexts;

c) Estimate the overall effectiveness of online patient simulations at improving clinical reasoning skills in undergraduate medical students;

d) Identify gaps in the literature to highlight where current evidence is lacking.

Thus, the current study can contribute towards informing the development and evaluation of future online patient simulations.

3.3 Methods

This systematic review was reported in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, Altman, & Group, 2009).

3.3.1 Inclusion and exclusion criteria

I describe the inclusion and exclusion criteria in **Table 3-1**. The criteria were developed based on previous reviews and my initial scoping searches of the literature, which helped to identify gaps in the current evidence base. I refined the criteria in consultation with my supervisors. I used the population, intervention, comparator, outcome (PICO) framework to identify the key concepts to include in my search (Schardt, Adams, Owens, Keitz, & Fontelo, 2007). I only included quantitative papers that attempted to assess the effectiveness of online patient simulations, as one of the main aims of the review was to generate an estimate of effectiveness from published studies

3.3.2 Search strategy

In July 2016 I applied a search strategy for the following databases: MEDLINE, EMBASE, CINAHL, ERIC, Scopus, Web of Science and PsycINFO. Further articles were identified by hand searching the reference lists of relevant papers. Search terms included medical students OR undergraduate medical students OR medical education AND computer-assisted instruction OR teaching OR online learning/education/teaching OR simulat* OR virtual realit* OR patient/virtual simulat* OR elearning OR electronic learning OR technology enhanced learning AND clinical decision making OR decision making OR clinical/diagnostic reasoning OR clinical judg?ment OR critical thinking/ reasoning. The full search strategy used in MEDLINE is included in **Appendix 1**.

76

Table 3-1 inclusion and exclusion criteria for my systematic review

Key Concepts	Criteria
Population	Undergraduate medical students.
	Excluded: health professionals, postgraduate students, other health students.
Intervention	Interventions that describe an educational method that is distributed and facilitated online and simulates a real-life clinical scenario between a 'physician' and 'patient. The student should emulate the role of a clinician by gathering data from the patient, interpreting information, and making diagnostic decisions (AAMC, 2007).
	Excluded: high fidelity simulators, manikins, standardised patients and decision support tools.
Comparator	No formal instruction, an alternative instructional method e.g. face-to-face or paper-based instruction.
Study type(s)	RCTs, crossover trials, quasi-experimental studies and observational studies.
	Excluded: cross sectional studies and qualitative designs.
Publication type(s)	Peer reviewed publications including theses.
	Excluded: conference papers, editorials letters, notes, comments and meeting abstracts. Studies not in English.
Outcome	Clinical reasoning skills are the thought processes required to identify likely diagnoses, formulate appropriate questions and reach clinical decisions (Higgs et al., 2008). Interventions that provided sufficient detail to establish whether it improved clinical reasoning skills in a written, oral or practical test. Commonly used synonyms for clinical reasoning were accepted e.g. clinical decision-making, clinical reasoning, problem-solving, critical thinking, clinical judgement skills.
Time	Publications from the year 1990, as this was when online learning was beginning to be described (Cook et al., 2008).

3.3.3 Study selection

I screened all the articles retrieved from the search by title and abstract for eligibility of inclusion. My supervisor (APK) double screened a randomly selected proportion of the abstracts to check for agreement and understanding of the eligibility criteria (4%, n = 100/2341). There was moderate agreement between reviewers and the criteria were reviewed, Cohen's Kappa = 0.58 (McHugh, 2012; Viera & Garrett, 2005). I resolved discrepancies in a consensus meeting and included studies if the abstract lacked enough detail to confirm eligibility. I screened all the full text articles and APK double screened a proportion of these articles (36%, n = 39/108), with moderate agreement (Cohen's Kappa = 0.52). Double screened articles were selected based on my uncertainty of whether to include them. I resolved any discrepancies in another consensus meeting. The discrepancies were due to poor descriptions of the interventions and outcome measures. In several studies it was difficult to establish whether the interventions used simulation, in particular whether students were required to gather information from the patient – these studies were excluded due to insufficient data. It was also unclear whether studies were measuring clinical reasoning skills according to my definition. My definition of clinical reasoning was the thought processes required to identify likely diagnoses, formulate appropriate questions and reach clinical decisions (Higgs et al., 2008). I included articles for full text review if there was doubt over their eligibility for inclusion.

3.3.4 Data extraction

I developed a data extraction form in Excel based on examples from previous reviews and added specific fields of interest relevant to my review. Data on study design, population, setting, intervention features, outcomes, results and limitations was extracted. APK and a GP registrar (SM) piloted the data extraction form by using it to extract data from one study each. Amendments were made to clarify the detail of data needed to be extracted for each field. I extracted data from eight studies included in the review, APK extracted data from three and SM extracted data from one. APK checked my data extraction and I double-checked the data extraction of APK and SM; we discussed any discrepancies.

3.3.5 Quality assessment

APK, SM and I assessed the quality of the studies in conjunction with data extraction. RP designed the quality checklist by incorporating items from two checklists: the medical education research study quality Instrument (MERSQI) and an adapted form of the Newcastle-Ottawa Scale (NOS), which have been used in combination with each other in previous reviews in this area (Cook et al., 2011; Cook et al., 2008; Reed et al., 2007; Wells GA, December 8 2016). I chose to incorporate two checklists as the NOS helped me to identify aspects of quality related to potential biases in the study design and sample selection and the MERSQI allowed me to identify other aspects of quality that are important to capture in medical education studies, such as the validity and reliability of outcome measures selected. Studies could receive a score of up to

14, with scores ranging from zero to four suggesting low quality, scores of five to nine suggesting moderate quality and scores of 10 to 14 indicating high quality.

3.3.6 Data analyses

For this review, I split analyses into two groups: online patient simulations vs no formal instruction and online patient simulation vs alternative methods of instruction. I described the intervention features, outcome measures and quality of studies by these groups to check for patterns in the data. I also conducted two separate meta-analyses to explore the impact of online patient simulations on clinical reasoning skills over nothing and in comparison to other teaching methods. It is important to acknowledge that students in the studies that compared online patient simulations to no formal instruction would have received implicit training in clinical reasoning skills in their curriculum, most likely through patient contact and clinical placements. Those studies that compared simulations to alternative methods of instruction, explicitly described in their study that they used an alternative method of delivering the same information as in the simulation, such as via a small group discussion.

I conducted a narrative synthesis of the studies to address the first two objectives of my review, which were to identify the intervention features and outcome measures studies used to evaluate their interventions. I aimed to identify: the theoretical basis of the interventions; the inclusion of key pedagogical techniques, such as feedback and reflection; whether it was a

group activity or individual; number of patient cases; clinical topic and duration of the intervention (Posel et al., 2009; Posel et al., 2015). I conducted two meta-analyses, including only studies that had a randomised design, to estimate the effectiveness of online patient simulations on clinical reasoning skills outcomes. While randomised designs traditionally determine efficacy rather than effectiveness (as they are usually conducted in artificially constrained settings), the studies included in this review were undertaken in real-world settings and so assessed effectiveness. My first analysis compared online patient simulations to no formal instruction, in which the simulation provided additional content to students. My second analysis compared online patient simulations to alternative methods of instruction, such as small group discussions of the same patient cases (Higgins & Green, 2005). I converted means of clinical reasoning outcome measures, as well as their standard deviations, partial eta squared scores and odds ratios, to a standardised mean difference (Hedges *g*) (Cochrane, 2011). For RCT designs I used the reported post-test means, standard deviations or odds ratios for the control and intervention groups in each study. For crossover studies, I used means adjusted for repeated measures or pooled the means across each intervention (Cochrane, 2011). For studies that had multiple assessments, I reported the average effect size across assessments. I considered effect sizes of 0.2 small, 0.5 as medium and those ≥ 0.8 large. The I^2 statistic was used to estimate heterogeneity across studies.

I found I^2 values larger than 50%, which indicated that there was substantial heterogeneity across studies (Green & Higgins, 2005). Therefore, the DerSimonian-Laird random-effects model was chosen to estimate the

pooled effect sizes (DerSimonian & Laird, 1986; Kontopantelis & Reeves, 2017). Given the small number of studies included, no sub-group analyses were performed, as the power to detect an effect would have been too low (Cochrane, 2011). Funnel plots were carried out to test for publication bias, but no significance tests were carried out due to the low number of studies (Egger, Smith, Schneider, & Minder, 1997; Sterne & Harbord, 2004). Sensitivity analyses were conducted to explore the nature of the heterogeneity across studies by removal of lower quality studies. Most effect sizes that were outliers were from studies that had lower methodological quality; thus, it was predicted that this accounted for most of the heterogeneity across studies. I excluded all studies that were rated as low quality (≤ 4) and those that scored on the lower end of the moderate scale (≤ 6) from both meta-analyses. All analyses were undertaken using statistical analyses package STATA version 15 (StataCorp, 2017).

3.4 Results

3.4.1 Study characteristics

The search strategy identified 5,461 records, of which 12 were included in the review. See **Figure 3-1** for the PRISMA flow diagram of the number of studies included at each stage of the review. Several studies (n = 96) were excluded due to insufficient detail in the study about whether their intervention fitted with my definition of online patient simulation. A table presenting the general characteristics of the included studies can be found in **Table 3-2.**

Figure 3-1 Flowchart explaining the study inclusion process
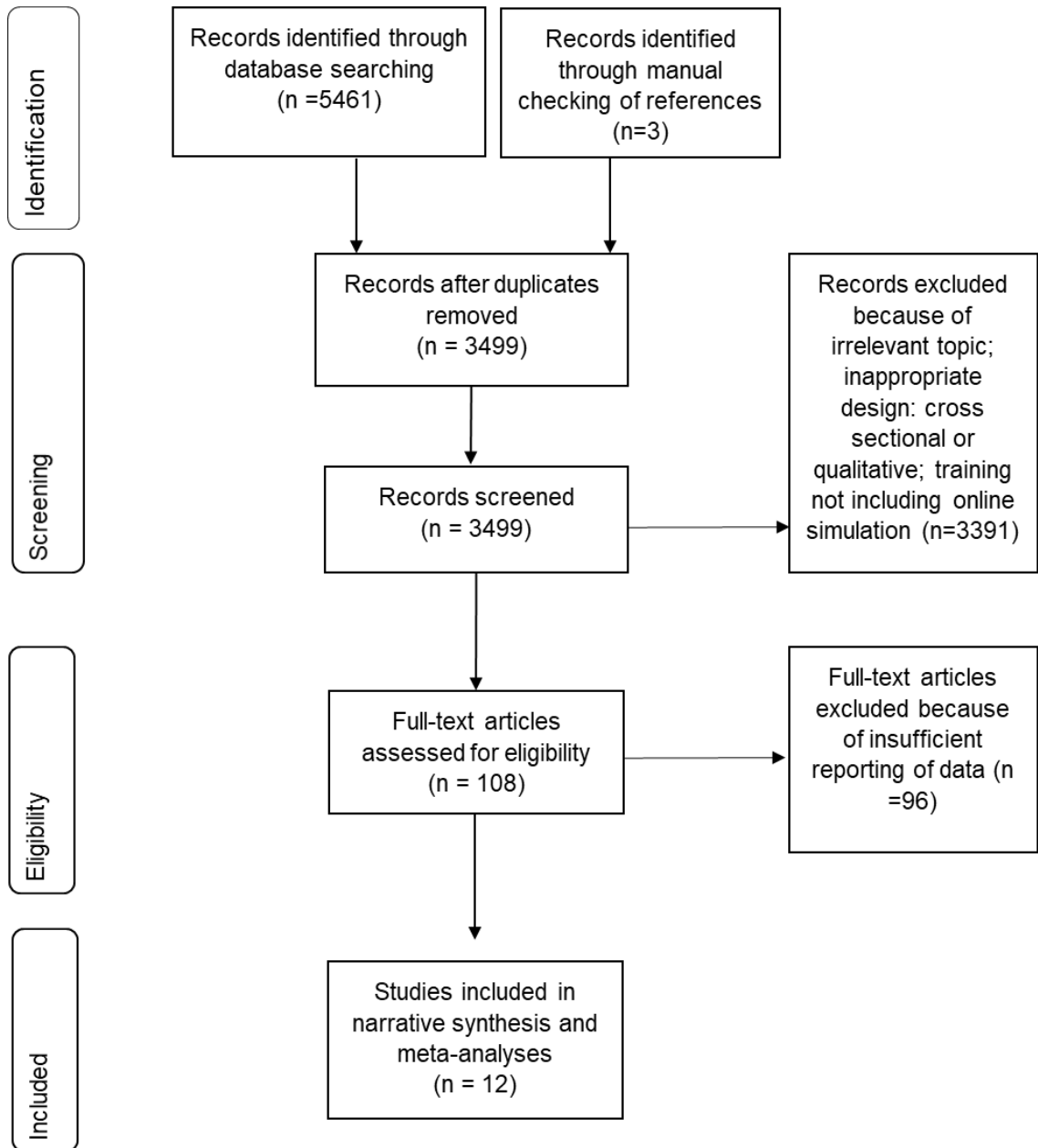
The most common study locations were Germany (33%, n = 4/12) and the USA (33%, n = 4/12). None of the studies included data from more than one medical school. Just over half of the studies were RCTs, (58%, n = 7/12). One was a non-randomised controlled trial, two were randomised crossover trials

and two used a single group pre-test and post-test design. The participants ranged from all years of undergraduate medical education, with three studies including students from more than one academic year. The number of participants in the studies ranged from 28-143 and a total of 853 students participated.

Just over half of the studies (n = 7/12, 58%) compared online patient simulations with alternative methods of instruction. The alternative methods of instruction included face-to-face learning methods, such as tutorials, small group discussions and lectures (n = 5/12, 42%); paper-based learning methods, which included text-based patient cases (n = 2/12, 17%) and real patient examination (n = 1/12, 8%; this was used in conjunction with face-to-face teaching). There was little detail across studies of how the information presented in the simulations was translated into different formats.

Table 3-2 Characteristics of included studies

Authors and reference number	Aim(s) of the study	Research Design	Number of participants (intervention and control group)	Uptake from eligible population	Outcome	Comparator	Outcome measure (V)[1]	Main results	Quality score (rating)[2]
Aghili et al. 2012[38]	To evaluate whether virtual patient simulation application improves clinical reasoning skills of medical students.	RCT	52 (29 IG, 23 CG)	100% of students on endocrinology course.	Clinical reasoning skills.	No formal instruction - traditional educational programs included didactic lectures, case-based small group discussions, bed-side face to face teaching sessions, and interactive teaching clinic.[3]	Diagnostic test (written test)	↑ Intervention produced significantly better scores on the diagnostic test compared to no formal instruction.	6 (Moderate)

Kalet et al. 2007 [28]	To assess the impact of individual WISE-MD modules on (1) short-term knowledge gain and students' satisfaction, and (2) clinical reasoning skills.	RCT	96 (52 IG, 44 CG)	100% of surgery clerkship students	Clinical reasoning	No formal instruction - the control group was given access to the module only during in the second half of the rotation. These students received no encouragement to view the module.	Script concordance test (V)	↑ Intervention produced significantly greater improvement in clinical reasoning compared to no formal instruction.	8 (Moderate)
Kleinart et al. 2015 [43]	To perform a validation of this novel educational approach and to examine whether the use of ALICE has positive impact on clinical reasoning and is a suitable tool for supporting the clinical teacher.	Single group pre-test–post-test comparison	62 (N/A)	100% of those attending surgical seminar	Clinical reasoning	N/A.	Patient cases	↑ Clinical reasoning significantly improved over time.	3 (Low)
Lehman et al. 2015 [30]	Investigated the effect of Virtual Patients combined with standard simulation-based training on the acquisition of clinical decision-making skills and procedural knowledge, objective skill performance, and self-assessment.	RCT	57 (30 IG, 27 CG)	11.9% of those in their 3rd or 4th year	Clinical decision-making skills	No formal instruction - they received the handout the same as the intervention group but no simulation.	Key feature test (V)	↑ Intervention produced significantly better clinical decision-making in the key feature test compared to no formal instruction.	12 (High)

86

Wu et al. 2014 [44]	To examine: 1) how a computer-based cognitive representation approach can be designed and implemented for learning clinical reasoning, and 2) the effectiveness of the approach in supporting the learning of clinical reasoning.	Single group pre-test–post-test comparison	50 (N/A)	58% of students on a residential course	Clinical reasoning	N/A	Concept maps (V)	↑ Problem solving significantly improved over time.	4 (Low)
Basu Roy & McMahon 2012 [41]	To explore video-based cases comparative impact on students' critical thinking; to evaluate student and tutor opinions and preferences on the use of video and text in case presentation.	Randomised crossover trial	28 (14 IG, 14 CG)	100% from 2 tutor groups	Critical thinking	Alternative method of instruction[4]. – paper-based. Written cases used a transcript of the video cases and access to the images and additional data for each case.	Students' discussions of patient cases (V)	↓ Intervention produced significantly lower odds of deep thinking compared to text-based cases.	8 (Moderate)
Botezatu et al. 2010 [39]	To explore possible superior retention results with Virtual Patients versus regular learning activities, by measuring the differences between early and delayed assessment results.	RCT	49 (25 IG, 24 CG)	100% from first clinical rotation in internal medicine	Clinical reasoning process	Alternative method of instruction – face-to-face, traditional teaching methods (lectures and small-group discussions covering the same topics).	Virtual patient cases (V)	↑ Intervention produced significantly better scores when solving the virtual patients, compared to those who received lectures and small group discussions on clinical cases.	6 (Moderate)

Devitt & Palmer 1998 [40]	To evaluate the intervention program by assessing whether it expanded students' knowledge base, improving data-handling abilities and clinical problem-solving skills.	RCT	71 (46 IG, 25 CG)	85% of year group	Clinical problem-solving skills	Alternative method of instruction – face-to-face. A comprehensive 90 min tutorial. It had the same material as the intervention in terms of text. There was a group discussion at the end of the week.	Multi-step clinical problem (patient case) (V)	_ Intervention produced non-significantly better scores clinical problem solving compared to a tutorial on the same topic.	6 (Moderate)
Kahl et al. 2010 [27]	To explore whether the addition of a systematic training in iterative hypothesis testing may add to the quality of the psychiatry course taught to fifth year medical students.	RCT	72 (36 IG, 36 CG)	Unknown	Diagnostic skills	Alternative method of instruction-paper-based and real patient examination. Problem-based learning, including paper cases of the same cases used in the intervention and group discussion.	Standardised patient (actor) (V)	↑ Intervention produced significantly better diagnostic skills compared to using paper cases and examination of real patients.	6 (Moderate)
Kamin et al. 2003 [33]	To determine whether critical thinking differs among groups receiving the same case with the same facilitator in one of three formats.	Non-RCT	65 students (25 IG-virtual, 20 – IG video, 20 – CG - text)[5]	100% of students on paediatric clerkship	Critical thinking	Alternative method of instruction – face-to-face and paper based. Face-to-face video modality, the students viewed the patients' story and discussed the case with the facilitator. Text modality allowed the patients' story to unfold in a narrative format and discussion of the case with a facilitator.	Students' discussions of patient cases (V)	↑ Intervention produced significantly better critical thinking than the text-based cases and face-to-face discussion of cases.	4 (Low)

88

McCoy 2014 [42]	This study investigates the utility of Virtual Patients for increasing medical student clinical reasoning skills, collaboration, and engagement.	Randomised crossover design	108 (54 IG, 54 CG)	100% of 1 class	Clinical reasoning skills	Alternative method of instruction – face-to-face. The tutor or the students commanded the mouse while discussing the same case on a PowerPoint projected screen.	Diagnostics competency task (written test) (V)	↓ Intervention significantly lowered learning gains compared to a PowerPoint facilitated small group discussion on the same topic.	8 (Moderate)
Raupach et al. 2009 [29]	To explore whether students completing a web based collaborative teaching module show higher performance in a test aimed at clinical reasoning skills than students discussing the same clinical case in a traditional teaching session.	RCT	143 (72 IG, 71 CG)	Unknown	Clinical reasoning skills	Alternative method of instruction – face-to-face. Small group 2-hour session was devoted to discussing the same case that was offered to the online group.	Key feature test (V)	– Intervention did not significantly improve clinical reasoning compared to small group discussions.	9 (Moderate)

---

[1] (V), the study demonstrated the validity or reliability of their outcome by either reporting inter-rater reliability, relationships to other variables, internal structure or content validity

[2] Scores of 0-4 equate to a low quality study, 5-9 moderate quality and 10-14 high quality

[3] No formal instruction is defined as teaching that is part of the curriculum on a particular topic. Online patient simulations in these studies were provided in addition to usual teaching and provided additional content to students which they would otherwise not have received.

[4] Alternative method of instruction is defined as an alternative way of delivering the same content as the online patient simulations but in a different way e.g. text based cases, small group discussions.

[5] Study did not report clearly how many participants were in each group but this was estimated based on information in the article.

3.4.2 Quality of included studies

Table 3-3 shows the quality of included studies**.** The average study quality was moderate ($M = 6.67$, $SD = 2.50$). Only one study was high quality, most were of moderate quality ($n = 8/12$) and three were of low quality. In all studies, it was unclear whether the sample represented the average learner in that community, as they did not report the proportion of the source population from which the patients were derived or demonstrate that the distribution of the main confounding factors was the same in the study sample and the source population. No studies reported they selected students from more than one medical school.

Table 3-3 Quality assessment of included studies

Authors	Randomised study design	Validity and reliability of outcome[1]	Sample from more than 1 medical school	Representativeness of cohort[2]	Selection of the non-exposed cohort[3]	Comparability of cohorts[4]	Blinding of outcome assessment	Incomplete outcome data[5]	Appropriate analysis	Power calculation	Total score (14)	Quality rating[6]
Aghili et al. 2012	1	0	0	0	1	1	1	1	1	0	6	Moderate
Kalet et al. 2007	1	3	0	0	1	1	0	1	1	0	8	Moderate
Kleinart et al. 2015	0	0	0	0	1	0	0	1	1	0	3	Low
Lehman et al. 2015	1	4	0	0	1	2	1	1	1	1	12	High
Wu et al. 2014	0	0	0	0	1	0	1	1	1	0	4	Low
Basu Roy & McMahon 2012	1	1	0	0	1	1	1	1	1	1	8	Moderate

Botezatu et al. 2010	1	1	0	0	1	1	0	1	1	0	6	Moderate
Devitt & Palmer 1998	1	0	0	0	1	2	0	1	1	0	6	Moderate
Kahl et al. 2010	1	1	0	0	1	1	0	1	1	0	6	Moderate
Kamin et al. 2003	0	1	0	0	1	0	1	0	1	0	4	Low
McCoy 2014	1	2	0	0	1	1	0	1	1	1	8	Moderate
Raupach et al. 2009	1	3	0	0	1	1	0	1	1	1	9	Moderate

---

[1] One point was given if the study reported the interrater reliability, content validity, internal structure and relationships to other variables of the outcome variable.

[2] One point was given if the study showed if the sample was representative of the average learner in this community.

[3] One point was given if they showed the comparison group was drawn from the same community as the exposed cohort.

[4] For randomised studies 1 point was given for randomisation and another for allocation concealment. For non-randomised studies one point was given if they controlled for baseline characteristics or outcome at baseline and two points were if awarded if they controlled for both baseline characteristics and the baseline outcome.

[5] One point was given if ≥75% of participants were included at follow-up.

[6] Scores of 0-4 equate to a low quality study, 5-9 moderate quality and 10-14 high quality.

3.4.3 Intervention features

Intervention features were described in **Table 3-4** and comparisons of these features between studies that compared online patient simulations to no formal instruction and alternative methods of instruction are described in this section.

3.4.3.1 Theories used to inform design

3.4.3.1.1 Studies that compared to no formal instruction

Two studies in this group provided a clear description of how theory informed the design of their simulations. Wu, Wang, Johnson, and Grotzer (2014) used cognitive learning theory to help them build an online simulation that facilitated the problem solving process through construction of concept maps. Concept maps were thought to help students to externalise their thought processes and allow them to visualise and make links between concepts. Another study described how the intervention was designed using the cognitive apprenticeship framework (Kalet, Coady, Hopkins, Hochberg, & Riles, 2007). This is a learning theory that recommends explicitly explaining experts' cognitive processes when solving complex tasks (Collins, Brown, & Newman, 1988). This was done by having expert clinicians demonstrate their thought processes for a patient case using a think aloud approach before students completed the task. Aghili et al. (2012) used theory to justify that using virtual patients facilitates active learning, which is an important principle in self-directed learning theory. Kleinert et al. (2015) justified their choice of using online patient

simulations by explaining how students can learn through trial and error. One study did not use theory to justify their choice of online patient simulation and its features (Lehmann et al., 2015).

3.4.3.1.2 Studies that compared to alternative methods of instruction

Problem-based learning (PBL) was the most commonly reported theory, which three studies described (Kamin, O'Sullivan, Deterding, & Younger, 2003; McCoy, 2014; Raupach et al., 2009). PBL was incorporated into the interventions by using the online patient simulations with groups of students, who discussed the clinical problems and worked collaboratively (Norman & Schmidt, 1992). McCoy (2014) also described using other learning theories, such as constructivism, deliberate practice, scaffolding and cognitive learning theory, to select features such as multiple patient cases and guidance from the tutor throughout the case. Another study used iterative hypothesis testing to design how students would elicit patient information by asking them to justify their choices throughout the patient consultation (Kahl et al., 2010). One study used a cognitive theory of multimedia learning to design their simulation (Basu Roy & McMahon, 2012; Mayer, 2009). The authors ensured that their simulation was designed in a way to minimise cognitive load by activating both visual and auditory channels of processing. Botezatu, Hult, Tessma, and Fors (2010) referred to using online patient simulations to improve knowledge retention, by using them as a way to practise and apply knowledge and used a formative assessment. One study did not mention theory (Devitt & Palmer, 1998).

Table 3-4 Further characteristics of the interventions and evaluations

Authors	Mode of delivery	Topic	No. cases	Maximum duration of intervention in hours	Did theory inform the use or design of the intervention?	Reflection used	Feedback used
Aghili et al. 2012	Individual	Endocrinology	2	Not reported	Unclear description	No	Yes
Kalet et al. 2007	Individual	Oesophageal cancer	3	1	Clear description	No	No
Kleinart et al. 2015	Group	Paediatric	1	Not reported	Clear description	No	Yes
Lehman et al. 2015	Individual	Paediatric basic life support	2	1	No theory used	No	Yes
Wu et al. 2014	Individual	Nephrology	5	20	Clear description	No	No
Basu Roy & McMahon 2012	Group	Human endocrinology and reproduction	2	4	Clear description	No	No
Botezatu et al. 2010	Individual	Haematology and cardiology	6	6	Unclear description	No	Yes
Devitt & Palmer 1998	Individual	Biliary disease	5	1.5	No theory used	No	Yes
Kahl et al. 2010	Group	Psychiatry	Not reported	70	Clear description	No but process of justifying hypotheses throughout the case may enhance reflection in action.	No
Kamin et al. 2003	Group	Surgical - carotid module and cholecystitis	2	Not reported	Clear description	No	Yes
McCoy 2014	Group	Heart palpitations	2	2.5	Clear description	No	Yes
Raupach et al. 2009	Group	Cardio-respiratory	1	10	Clear description	No	No

3.4.3.2 Key pedagogical features: Feedback and reflection

3.4.3.2.1 Studies that compared to no formal instruction

Feedback was provided in a variety of ways. Kleinert et al. (2015) used a virtual instructor to summarise the underlying declarative knowledge, optimal procedural pattern and showed a video demonstration of the surgical procedure. One study gave feedback by comparing their decision with the expert decision (Lehmann et al., 2015). Aghili et al. (2012) used their online programme to inform students about questions, examinations, laboratory tests or requested images that were unnecessary and others that should be selected. Two studies did not provide information on how feedback was given (Kalet et al., 2007; Wu et al., 2014) and none reported how reflection was facilitated.

3.4.3.2.2 Studies that compared to alternative methods of instruction

One study had a feedback module that had a detailed case discussion and actual patient follow-up (Botezatu et al., 2010). Another study gave students a score to compare to other users (anonymously) and they could print out a critique of the case (Devitt & Palmer, 1998). McCoy (2014) used tutors to debrief students on the 'muddy points' in the case and provided immediate prefabricated feedback on students' decisions from a virtual tutor throughout the patient cases. Kamin et al. (2003) reported that the tutor provided a group assessment but gave few details. Two studies did not provide information on how feedback was given (Basu Roy & McMahon, 2012; Raupach et al., 2009). Reflection was only implied in one study that used the iterative hypothesis

testing model to design their simulation (Kassier, 1983). This method allowed students to pause and reflect continually throughout the consultation.

3.4.3.3 Other pedagogical features

3.4.3.3.1 Studies that compared to no formal instruction

All studies were delivered to individuals rather than groups, except one (Kleinert et al., 2015). Topics varied, but two studies created paediatric based patient cases (Kleinert et al., 2015; Lehmann et al., 2015). The number of cases in the intervention ranged from one to five; the most common number of cases was two. In the three studies for which the duration of the intervention was reported, the duration ranged from 1-20 hours (Kalet et al., 2007; Lehmann et al., 2015; Wu et al., 2014).

3.4.3.3.2 Studies that compared to alternative methods of instruction

Most studies were delivered to groups rather than individuals, except for two (Botezatu et al., 2010; Devitt & Palmer, 1998). Clinical topics varied, but cardiology topics were most common (Botezatu et al., 2010; McCoy, 2014; Raupach et al., 2009). The number of cases in the intervention ranged from one to six; the most common was two cases. The duration of the intervention of the duration ranged from one and a half, to 70 hours.

3.4.4 Outcome measures

3.4.4.1 Studies that compared to no formal instruction

Clinical reasoning was measured using a wide range of instruments. One study used a written test, which presented essay questions based on two patient cases, and asked the students to describe the essential points in history and physical examination (Aghili et al., 2012). One study used students' performance on an additional patient case, such as whether they got the correct diagnosis and selected a suitable therapy (Kleinert et al., 2015). Another study evaluated students' concept maps on five aspects of performance, including how students: observed critical information; formulated hypotheses; performed reasoning for justification and generated concepts and concept relationships (Wu et al., 2014). Two studies (Kalet et al., 2007; Lehmann et al., 2015) used methods of assessing clinical reasoning that have been previously validated: the key features problem and script concordance test (Charlin, Roy, Brailovsky, Goulet, & van der Vleuten, 2000; Page, Bordage, & Allen, 1995). Three studies validated their measures of clinical reasoning (Kalet et al., 2007; Lehmann et al., 2015; Wu et al., 2014). All studies reported a significant positive effect of the intervention using these outcome measures.

3.4.4.2 Studies that compared to alternative methods of instruction

Two studies assessed reasoning by assessing critical thinking and evidence of deep thinking in students' discussions (Basu Roy & McMahon, 2012; Kamin et al., 2003). Both studies used models of critical thinking, such as

Garrison's model, which allowed them to categorise students' discourse into one of five stages of critical thinking (Garrison, 1991). One study measured reasoning via a written diagnostic competency task, which measured whether students could make accurate diagnoses and management plans for patients using multiple-choice questions (MCQs; McCoy, 2014). Two studies used previously validated ways of measuring clinical reasoning (including a key features problem and a standardised patient) (Kahl et al., 2010; Raupach et al., 2009). Two studies assessed reasoning using a further patient case (Botezatu et al., 2010; Devitt & Palmer, 1998). All these studies validated the outcome measures they used in some way. Only one study reported a further follow-up assessment after the immediate post-test assessment to assess sustainability of the possible impact of the intervention (Botezatu et al., 2010).

3.4.5 Quality of included studies by comparator

3.4.5.1 *Studies that compared to no formal instruction*

Quality ranged widely in studies that compared online patient simulations to no formal instruction, with one study scoring a total of three and another 12. Two were classified as low quality studies with scores under five (Kleinert et al., 2015; Wu et al., 2014). This was mainly because these studies did not use a randomised study design and failed to validate their outcome measures. The study with the highest quality score did not sample from more than one medical school and did not demonstrate that the sample was representative of the population but otherwise was of high quality (Lehmann et al., 2015). Those that received moderate scores mostly failed to use multiple ways of ensuring the

validity and reliability of their outcome measure, failed to blind the outcome assessor to group allocation and failed to do a power calculation (Aghili et al., 2012; Kalet et al., 2007).

3.4.5.2 Studies that compared to alternative methods of instruction

Quality was slightly less varied in studies that compared online patient simulations to alternative methods of instruction, with one study scoring a total of four and another a total of nine. Only one was classified as low quality and this was mainly because it: did not use a randomised study design; failed to use multiple ways of ensuring reliability and validity of the outcome measure; did not control for baseline characteristics; had relatively incomplete outcome data and failed to do a power calculation (Kamin et al., 2003). All of the rest of the studies were rated as moderate in quality. Three studies were at the lower end of the moderate scale with a score of six (Botezatu et al., 2010; Devitt & Palmer, 1998; Kahl et al., 2010). Their scores were lower because they mostly failed to use multiple ways of ensuring validity and reliability of the outcome measure, failed to blind the outcome assessor to group allocation and failed to do a power calculation. The other three studies scored at the upper end of the scale, scoring eight to nine, because they addressed these issues (Basu Roy & McMahon, 2012; McCoy, 2014; Raupach et al., 2009).

3.4.6 Engagement

3.4.6.1 Studies that compared to no formal instruction

Most of the studies that compared online patient simulations to no formal instruction had 100% uptake. This was largely because students were recruited from a subsection of a student year group and the intervention for most of these cases appeared to be integrated into the module or curriculum. For example, Kleinert et al. (2015) only recruited students who attended a surgical seminar. Uptake was considerably poorer for those recruiting entire year groups and when participation appeared to be voluntary. Lehmann et al. (2015) only achieved 11.9% uptake across two year groups and Wu et al. (2014) achieved 58% across one year group. All studies had relatively good completion, with all studies reporting that more than 75% of their original students remained at follow-up.

3.4.6.2 Studies that compared to alternative methods of instruction

Similarly, most studies that compared online patient simulations to alternative methods of instruction had 100% uptake. Again, this appeared to be because they recruited students from small groups e.g. students on specific rotations and had integrated the evaluation and intervention into the module. One study that recruited from the year group on a voluntary basis achieved 85% uptake (Devitt & Palmer, 1998), but two studies did not report how many out of those eligible were actually recruited (Kahl et al., 2010; Raupach et al., 2009). All studies achieved good follow-up and retention rates as only one study had

less than 75% of participants remain in the study from baseline (Kamin et al., 2003).

3.4.7 Reported effectiveness

All studies that compared online patient simulations to no formal instruction showed a significant improvement. Two studies found online patient simulations had a significantly negative effect compared to alternative methods of instruction, such as face-to-face learning and paper-based learning (Basu Roy & McMahon, 2012; McCoy, 2014). Two studies found no significant difference in their intervention compared to face-to-face learning methods (Devitt & Palmer, 1998; Raupach et al., 2009). Three studies found significant positive effects of online patient simulations compared to paper-based, face-to-face and real patient examination instructional methods (Botezatu et al., 2010; Kahl et al., 2010; Kamin et al., 2003). One study found the significant positive effect of the intervention remained after 4.5 months compared to face-to-face learning (Botezatu et al., 2010).

3.4.8 Meta-analysis

Using only data from the studies that used randomised study designs, meta-analyses were conducted to estimate the effectiveness of the interventions on clinical reasoning skills. According to the meta-analyses (**Figure 3-2**) online patient simulations had a significant and substantial benefit on clinical reasoning skills, compared with no formal instruction. The pooled

effect size was 1.29 (95% CI, 0.25-2.33), $p \leq 0.05$. However, there was high

heterogeneity across studies (I^2 = 92.7%). The funnel plot was asymmetric,

which indicated that there might be publication bias. Online patient simulations

had a non-significant medium positive impact on clinical reasoning skills,

compared with alternative methods of instruction (**Figure 3-3**). The pooled

effect size for these interventions was 0.39 (95% CI, -0.17-0.95) p = 0.17. Again,

there was high heterogeneity across studies (I^2 = 94.71%) and the funnel plot

was asymmetric, which suggested possible publication bias.

Figure 3-2 Forest plot showing overall effect sizes of included interventions comparing online patient simulations to no formal instruction
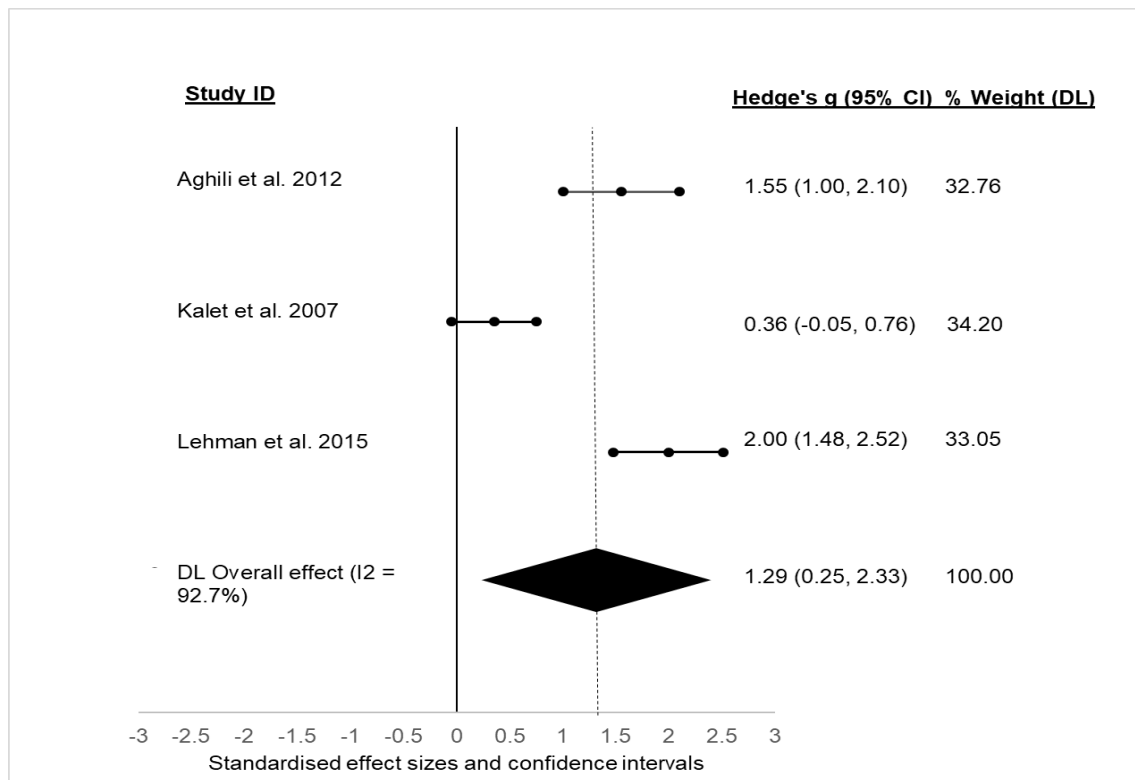
Figure 3-3 Forest plot showing overall effect sizes of included interventions comparing online patient simulations to no alternative methods of instruction
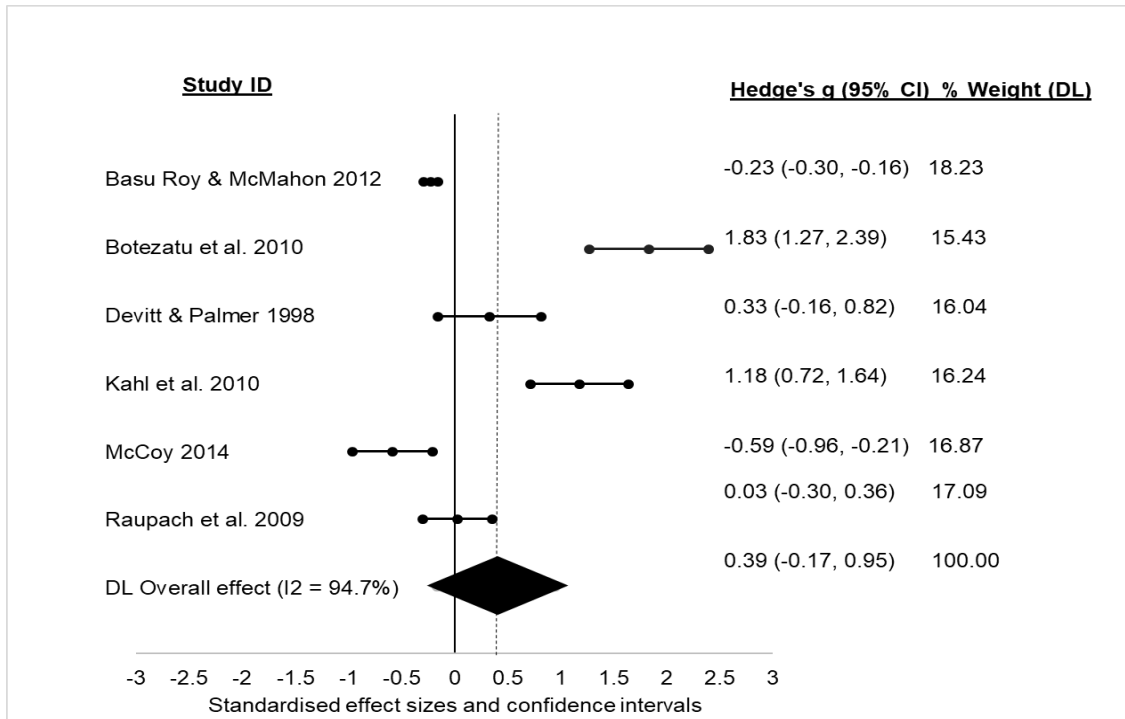
3.4.8.1 Sensitivity analysis

When comparing online patient simulations to alternative methods of instruction, it can be seen that studies of lower quality (quality score ≤ 6), regularly reported that online patient simulations had a small to large positive impact on clinical reasoning skills (effect sizes ranging from 0.33 to 1.83). Studies that were of higher quality (quality score ≥ 8) found a moderate negative impact, or no impact of the intervention on clinical reasoning skills (effect sizes ranging from -0.58 to 0.03). Therefore, a sensitivity analysis was conducted to decide on whether to exclude lower quality studies. These analyses found that online patient simulations had an even larger but seemingly non-significant effect on clinical reasoning skills, compared with no formal instruction. The

pooled effect size was 1.78 (95% CI, 1.35-2.20), $p = 0.25$. There was much less heterogeneity across the studies ($I^2 = 25.3\%$), suggesting that quality was a main contributor to the heterogeneity across these studies. However, only two studies were included in this analysis (Kalet et al., 2007; Lehmann et al., 2015), so it is difficult to draw reliable and valid conclusions about these results. Indeed, the model produced surprising results as the confidence intervals contained one, yet the significance value was not below 0.05. An imprecise estimate of the effect may have been caused by the lack of data in the model.

The sensitivity analysis comparing online patient simulations to alternative methods of instruction found online patient simulations had a small negative impact on clinical reasoning skills, which approached significance, compared with alternative methods of instruction. The pooled effect size for these interventions was -0.24 (95% CI,-0.49-0.01) $p = 0.06$ (Basu Roy & McMahon, 2012; McCoy, 2014; Raupach et al., 2009). However, there was still moderately high heterogeneity across studies ($I^2 = 65.3\%$). Further inspection of the intervention features suggested that the interventions were relatively similar for the three included studies, as they were delivered to small groups and had a similar number of cases. The main cause of heterogeneity was likely the difference in outcome measures used to assess clinical reasoning (as each used a different measure). In addition, two studies used a randomised crossover study design, whereas one used a parallel RCT design.

3.5 Discussion

I found some evidence that online patient simulations improve clinical reasoning in medical students compared to no formal instruction. However, when poor quality studies were excluded, positive effects from online patient simulations were non-significant. Online patient simulations appeared to have no significant benefit over alternative instructional methods, such as PBL. My narrative review identified some intervention features and outcome measures that could help future studies to build evidence-based interventions that could improve clinical reasoning skills.

3.5.1 Comparisons with other literature

With the caveat of the uncertainty of my conclusions from my meta-analyses, my results partially concur with previous reviews. Most reviews have found significant benefits to using internet-based learning platforms on general learning outcomes compared with no intervention (Consorti et al., 2012; Cook et al., 2010; Cook et al., 2008). While my initial analysis found a significant positive effect, my sensitivity analysis showed this benefit was not statistically significant when lower quality studies were removed. A previous review also found a negligible to small positive effect when comparing computerised instruction to alternative methods of instruction (Cook et al., 2010). My initial analysis concurred with this finding, but my sensitivity analysis found a small and non-significant negative effect of online patient simulations. Overall findings appear to suggest that providing some form of explicit clinical reasoning

teaching is likely to be better than no explicit teaching on clinical reasoning skills, but the instructional method used may not be important. However, so far conclusions have been drawn from low quality studies; more methodologically robust studies are needed to make more precise estimates of effectiveness.

Of course, effectiveness is just one consideration for choosing between instructional methods. Since previous reviews and this review suggest that online patient simulations are at least comparable to traditional methods of teaching, other factors, such as cost and availability of resources, might also influence medical educators decisions of how to teach skills (Cook et al., 2010; Cook et al., 2011; Cook et al., 2008). Educators may choose to use online patient simulations because there are substantial long-term practical benefits to using computer-based instructional methods over other teaching methods. For example, computer-based methods can be more easily and frequently updated, tailored to the learner's needs, provide immediate feedback and distributed to students who may be in remote areas (or at home) (Bradley, 2006; Issenberg et al., 2005; Vaona et al., 2015). With similar effectiveness, these practical advantages could make them a more viable teaching method than other choices (Vaona et al., 2015). However, the initial costs of creating such tools and concerns over technical issues and engagement with the online programmes may make them less feasible as an instructional method.

Educators also select instructional methods based on theory and the proposed added value of certain instructional methods. Of the included studies

that mentioned that their interventions or evaluations were based on theory, they referred to cognitive and learning theories as the underlying rationale for developing online patient simulations to improve clinical reasoning skills for students. The key mechanisms of change proposed in these studies were that online patient simulations offer an immersive experience that allows students to engage in a task from beginning to end. This enables students to be more engaged and active in their learning; this is a core principle of constructivist learning theories, such as self-directed learning theory and experiential learning theory (Kolb, 1984; Mann, 2011; Merriam et al., 2012; Rutherford-Hemming, 2012). Other studies in this review drew on the theory of deliberate practice that emphasises the importance of developing expertise through repetition, and explained how online patient simulations allow for repetitive practice and learning by trial and error (Ericsson, 2008). Studies based on cognitive theories recognised the opportunity for online patient simulations to provide students with reflective and metacognitive experiences with a wide range of patient cases, which helps students to develop their cognitive abilities and create complex mental representations of illnesses (Merriam et al., 2012; Rutherford-Hemming, 2012; Schön, 1983). However, none of the studies in my review described using specific models of reflection, such as the reflective practice model. Use of such models could have led to theoretically important intervention features, such as feedback and reflection, being more explicitly used and described (Mamede & Schmidt, 2004). While none of the studies aimed to test theory, more studies are needed that aim to do this to understand how, why and what works in such interventions (Cook, 2005; Cook et al., 2010). Indeed, no

studies included a logic model, which would have helped to elucidate the mechanisms of change.

Previous reviews in this area found studies tended to have poor descriptions of their intervention features (Cook et al., 2011). I found that most included studies reported that they used feedback, which previous research indicated is associated with positive learning outcomes, and has been recommended for inclusion in online patient simulations (Cook et al., 2010; Posel et al., 2015). My review attempted to understand how feedback was delivered to inform future studies. Similarly, to other reviews I found that details on how feedback was delivered were often brief. Studies mainly provided feedback by comparing students' performance to that of a clinician. Some used scores and compared these between students anonymously, which potentially increased competition and motivation to improve (Rutherford-Hemming, 2012). Most studies provided feedback after the patient cases but one provided feedback during the case, which may have provided more active guidance and scaffolding for students (Sharma & Hannafin, 2007).

It was surprising that few studies reported that they sought to prompt reflection, given that feedback and reflection are often used together, and are seen as a key component of several learning and cognitive theories (Posel et al., 2015). Interestingly, the only study that prompted students to reflect, encouraged students to reflect-in-action, i.e. reflect during the patient consultation rather than just at the end of the task (Kahl et al., 2010; Schön,

1983). This was achieved by utilising the iterative hypothesis testing approach, that requires students to continually justify their hypotheses (Kassier, 1983). A recent trial suggested that using this approach can improve clinical reasoning skills (Nendaz, Gut, Louis-Simonet, Perrier, & Vu, 2011). Furthermore, online patient simulations may be particularly suited to enabling reflection-in-action, as they allow students to pause their reasoning at key points in a patient consultation, which may be inappropriate or inconvenient to do with real patients.

Studies varied in whether they were delivered to groups or individuals and most had at least two patient cases and lasted at least an hour, which gives an indication of the minimum requirement for an online patient simulation. Most current online patient simulations appeared to focus on paediatrics and cardiology. More patient cases that focus on clinical reasoning skills in the context of general practice could be useful given the limited amount of clinical contact students have in general practice settings (McDonald, Jackson, Alberti, & Rosenthal, 2016). Additionally, cases that focus on life-threatening respiratory conditions, such as lung cancer, in general practice would also be useful, as research suggests clinicians are prone to clinical reasoning errors in this context (Sheringham et al., 2016).

In common with other reviews, there was a lack of validity and reliability in outcome measures used to assess clinical reasoning skills (Cook et al., 2010; Cook et al., 2011). Many studies did not use previously validated measures,

such as the key features problem, script concordance test and standardised patients, but developed their own measures. Most studies used some form of patient case to assess clinical reasoning skills, which included assessing their performance on gathering information, diagnosing patients and selecting appropriate tests. This approach may be necessary given clinical reasoning is not a generic skill and creation of patient cases similar to the intervention allows students to demonstrate skills in a similar context and of similar content to what they were taught via the simulations. Clinical reasoning skills are highly context and content dependent and there is unlikely to be one generic measure of clinical reasoning skills that could capture all of its complexity in a standardised way. While we should not expect the same measure to be used across all studies, studies should still seek to apply a common framework (such as that provided by the key features problem) and validate the measures used.

3.5.2 Strengths and limitations

Strengths of this review included the thorough search of the literature and the contribution to the understanding of what is currently known about the effectiveness of online patient simulations on medical students' clinical reasoning skills. However, the small number of studies included and the low number of participants, meant there was low power to detect an effect in the meta-analyses, which limited this review. The meta-analyses showed substantial heterogeneity across studies. The substantial heterogeneity was likely caused by the varied content of the simulations, the different measures of clinical reasoning and the variation in the characteristics of samples. I also

found that studies with lower effect sizes tended to be of higher methodological quality than those with higher effect sizes. Indeed, funnel plots suggested that publication bias was present; suggesting lower quality studies that found positive findings were more likely to be published than those that found non-significant findings. Thus, one should be cautious when interpreting the results of the meta-analyses.

Researchers have identified the need for studies to elucidate what intervention features would be most effective for online patient simulations by comparing studies using the same medium (Cook, 2005; Cook et al., 2010; Cook et al., 2011). However, it was not possible to compare the effectiveness of different intervention features because I found no studies that compared intervention features within the same medium, such as a computer-based study that compared two different ways of delivering feedback. Researchers have argued that comparisons made across mediums (e.g. computer-based vs paper-based) are not particularly valuable given that it is not possible to disentangle the potential influences of multiple confounders that could have caused the effect (Cook, 2005). Nevertheless, because tools like online patient simulations are only in the infancy of development, comparisons to 'usual teaching' are still valuable, as they provide information on uptake and engagement with these novel ways of teaching. Indeed, this review showed that there was limited information available about engagement in these interventions. It is important to demonstrate first that online patient simulations can be implemented into the curriculum and are accepted by faculty and

students, before evaluations determining effectiveness and comparing features are undertaken.

The outcomes of the studies were also difficult to compare because most studies did not use previously validated measures of clinical reasoning, such as key features problems. Most studies (n = 10/12) did report validity, but most reported only one type of validity or reliability for their outcome. Without sufficient knowledge of the validity and reliability it was difficult to compare outcomes across studies and to create a benchmark for an acceptable measure of clinical reasoning for future studies. Ideally further research should report evidence of validity if they develop their own measures and validate these measures in the author's centre if possible.

None of the studies in my review included a logic model to explain how and why their intervention worked, and why they used specific outcome measures. Logic models are useful in improvement interventions because they ensure that aspects of the intervention, and the mechanisms through which the intervention are thought to lead to intended outcomes, are clear. They can also provide a framework for the evaluation (Davidoff, Dixon-Woods, Leviton, & Michie, 2015). The lack of clarity on these factors in the current literature indicated to me that a logic model would benefit the development of an online patient simulation learning tool, so that these factors are clear to future researchers and users. The review itself was less informative than I had expected, for the development of my own logic model to show how online patient simulations may

be effective. There was a lack of high methodological quality RCTs in the field to establish effectiveness and a lack of detail on the theoretical understanding of how and why the interventions might work. It may have been informative for my logic model, to have included qualitative studies in the review that sought to elucidate candidate explanations for why and how online simulations lead to impacts (or not). Inclusion of such studies would also have aided my understanding of the wider use and design of online patient simulations. However, a broader review was beyond the scope of this PhD and the aims of this review. The narrative review of the quantitative studies in this review did provide some insights into the theoretical understanding of how online patient simulations work. The broader literature, including qualitative studies, were reflected on in the introduction and discussion of the thesis and in this chapter, which contributed to the logic model described in **Chapter 4**.

The review was also limited by the small percentage of studies that were double screened for inclusion and the moderate inter-rater agreement (McHugh, 2012; Viera & Garrett, 2005). The low number of abstracts and full-text articles that were double screened was due to the time constraints of my PhD and the capacity of the other reviewers. The moderate agreement was due to the lack of detailed information in the studies about what their interventions were, who their populations were, what their outcome measures were and the wide range of terms that can be used to describe online patient simulations and clinical reasoning skills. All discrepancies were resolved in discussion meetings,

and any studies that I was unsure of were double screened in the full-text review.

3.6 Summary

This chapter addressed the first research question of my thesis; to explore what is currently known about the effectiveness of online patient simulations on the clinical reasoning skills of medical students. My review indicated that online patient simulations may be effective at teaching clinical reasoning skills to medical students compared to no formal instruction but effectiveness was not significant after the removal of low quality studies. It highlighted that published studies have poor descriptions of intervention components and added little to the understanding of the underlying rationale for using online patient simulations to improve clinical reasoning skills. I observed a particular weakness in the application of theory in current online patient simulations, as even those studies that reported using theory, did not include features that cognitive and learning theories have recommended, such as reflection (Croskerry, 2003; Issenberg et al., 2005). Furthermore, there were no online patient simulations that covered respiratory problems encountered in primary care, despite research suggesting that this is a context in which many clinical reasoning errors occur (O'dowd et al., 2015; Sheringham et al., 2016). These findings suggest that for a newly developed tool it should be clear how that intervention is based on theory and a logic model should be developed to show how the proposed mechanisms of change can be observed and measured

and lead to a robust evaluation. Therefore, a logic model that explains how online patient simulations could improve students' reasoning skills will be described in **Chapter 4**.

I also found that there were relatively few evaluations of such tools of sufficient quality, and, therefore, there is a need for more robust evaluation studies. This review showed there was limited information from current studies on uptake and dropout rates, recruitment methods and outcome measures. Therefore, it would be useful to conduct a feasibility RCT to first understand feasibility outcomes, such as uptake and retention, before a definitive RCT should be used to assess effectiveness. An evaluation is also required to understand how online patient simulations help students to learn these skills and develop the logic model. My review has also shown me different ways of measuring clinical reasoning skills in an evaluation. It has reinforced the importance of validating outcome measures to understand the impact of an intervention, as this was noted as a particular weakness in previous studies. It has shown how it is possible to use online patient simulations not just for training but for enhancing our understanding of the clinical reasoning process. By looking at the activities of students, such as how students gather information or reasons why they chose diagnoses, it is possible to make visible the thought processes that underlie students' clinical decisions.

4   Development of an online patient simulation

In this chapter, I propose a logic model by synthesising learning from the theoretical and empirical literature from **Chapters 1 and 3**. I then describe a new simulation tool to which I contributed.

4.1   My contribution towards the development of the tool

This PhD contributes towards a wider project, funded by the National Institute for Health Research (NIHR) Policy Research Unit (PRU) in Cancer Awareness, Screening and Early Diagnosis, to develop a tool to improve the clinical reasoning skills of medical students. My aim was to ensure this tool was based on empirical evidence and sound theory. To achieve this aim I needed to answer my second research question, to explore what the theoretical and empirical literature indicated about the required features of online patient simulations to promote medical students' clinical reasoning skills. My research objectives to achieve this aim were to:

a) Provide clarity on how online patient simulations could be effective in promoting medical students' clinical reasoning skills. I did this by developing a logic model that synthesised learning from the theoretical and empirical literature from **Chapters 1 and 3**.

b) Use this learning to inform the design of the PRU online learning tool. I describe the context for the PRU tool and the resulting tool to which I contributed.

4.2 Learning from theory

In **Chapter 1**, I summarised how cognitive theories, clinical reasoning theories and theories of learning suggested that approaching teaching clinical reasoning skills via simulation is beneficial to students. These theories suggest that simulations could be beneficial to developing analytical clinical reasoning skills if they incorporated feedback, scaffolding and prompts for reflection, to facilitate reflection and metacognition (Croskerry, Singhal, & Mamede, 2013; Ericsson et al., 1993; Mamede et al., 2008; Schön, 1983; Sharma & Hannafin, 2007). Simulations could also help students to have realistic experiences with a variety of patient cases by facilitating repeated exposure to clinical cases. This could aid the development of non-analytical reasoning and help students to build complex mental representations of problems. It could be more feasible to deliver a larger number of simulated patient cases online than through face-to-face interaction with simulated patients, as those methods could be more resource and time efficient for staff to organise and deliver in the long-term (IoM, 2015; Page et al., 2016). Furthermore, online patient simulations could allow for more iterative feedback and reflection during a patient case than through face-to-face interaction with simulated patients or real patient contact, where interrupting the flow of a consultation may not be feasible.

4.2.1 Learning from my theoretical model of medical students' clinical reasoning

In **Section 1.5** I described my theoretical model of students' clinical reasoning based on previous clinical reasoning models. In **Figure 4-1** I have highlighted in red the specific parts of the theoretical model that online patient simulations could effect and how these effects could be detected by the simulations. My theoretical model indicated that an online patient simulation could improve reasoning by influencing the thought processes or mental representations of medical students and could also alleviate the influences of cognitive biases that affect thought processes (Component B in **Figure 4-1**). If thought processes improve then this will have a positive impact on student behaviours (how they gather data – Component C) and on their consultation outcomes (their evolving diagnosis and management plan - Component D), which will feed back into their student attributes and help them to build complex mental representations of similar problems that they can take forward to their next case (Component A).

In order to influence thought processes an online patient simulation should aim to develop both analytical and non-analytical thinking (Component B). Online patient simulations could allow students to develop their non-analytical thinking by exposing them to a variety of clinical cases, which can help them develop complex associations between symptoms and diagnoses and, therefore, more complex mental representations. Online patient simulations could develop analytical

thinking skills by providing scaffolding and prompts for reflection during a patient case. These educational techniques could help students to engage in reflection and metacognition and provide them with a structure to guide their thinking when assessing a case. For example, it could guide them to consider alternative hypotheses or to consider what information is needed to confirm and refute their hypotheses; thereby, also mitigating the effect of cognitive biases that are known to influence reasoning. Online patient simulations could also provide guidance on what strategies are useful in particular contexts, as clinical reasoning is context dependent and students need to develop flexibility in adapting their reasoning for different contexts. Online patient simulations can also be used to detect and assess behaviours (data gathering), which reflect the underlying and unobservable thought processes that are being influenced (Component C). The approach students take to gathering data can be captured by the simulation by observing what questions students ask, their rationale for these questions and diagnoses, together with the order they ask the questions (Component C).

Figure 4-1 Theoretical model for medical students' clinical reasoning with points of intervention for online patient simulations and how they might be observed via online patient simulations
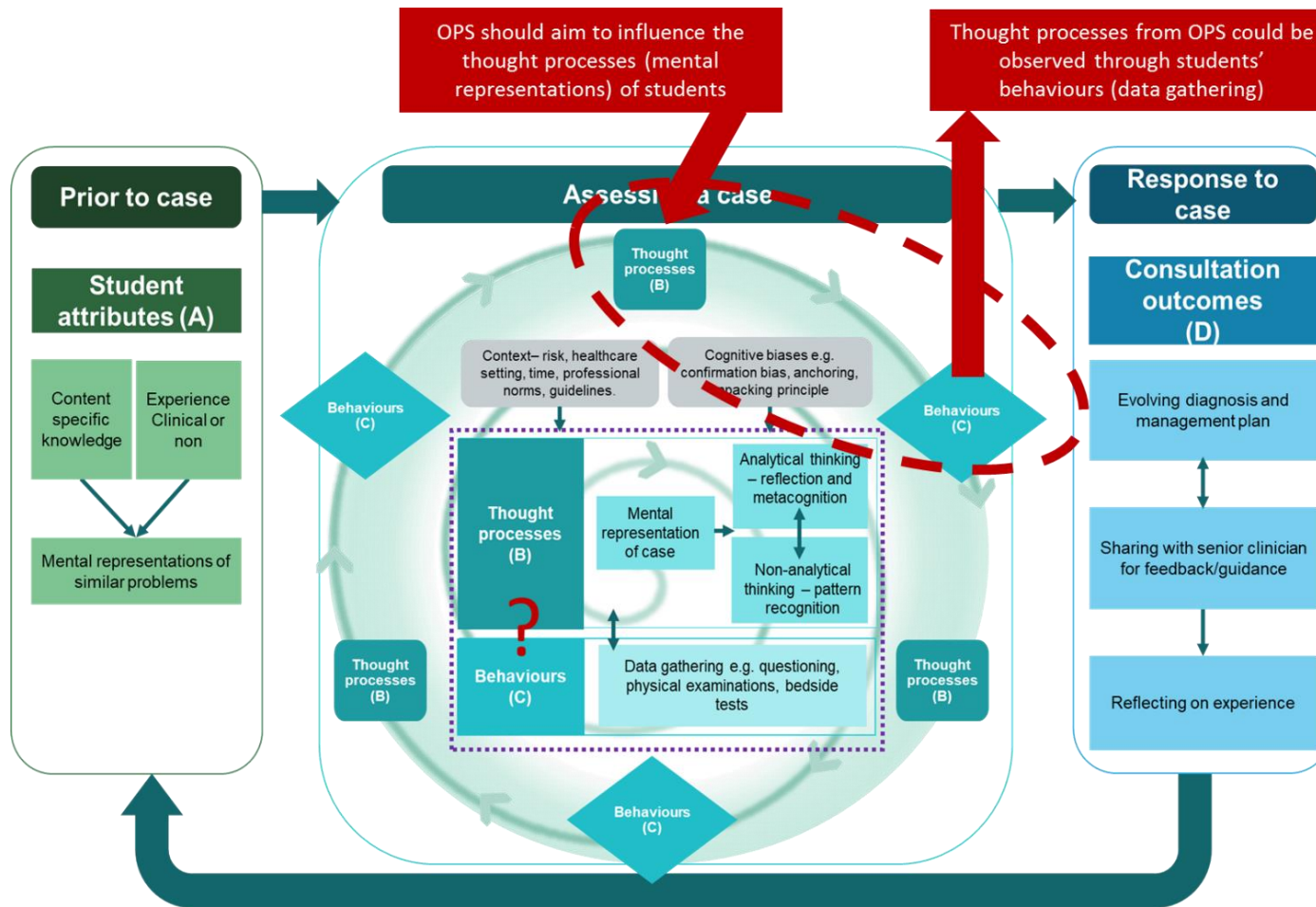
4.3 Learning from health services research

In **Chapter 1**, I also discussed the clinical and public health importance of improving clinical reasoning skills, to reduce missed opportunities for diagnosis. It may be particularly important to improve these skills in the primary care context, given that it is known that errors and delays in the early diagnosis of conditions like cancer in primary care can contribute to poor patient outcomes (Bradley et al., 2018; Rubin et al., 2015). Indeed, medical students may benefit more from simulations set in general practice, as they currently have little explicit training on clinical reasoning in these contexts (Harding et al., 2015; Higgs & Jones, 2000). One of the main reasons some cancers, such as lung cancer, and other conditions like sepsis can be missed in primary care is because patients often present with vague and non-specific symptoms (Kostopoulou et al., 2008; Rubin et al., 2015). I propose that using online patient simulations set in the context of primary care, with patients who have vague symptoms indicative of serious and life-threatening conditions like cancer, would be useful training for future clinicians. This training could help students to form suitable reasoning strategies to manage these patients more appropriately as future doctors. My theoretical model in **Chapter 1** showed that both non-analytical and analytical thinking and the testing of diagnostic hypotheses are needed to make an appropriate diagnosis. Online patient simulations could be particularly useful in helping students to develop their analytical thinking through scaffolding, refection and provision of feedback, which may be more appropriate than non-analytical thinking for clinical cases in primary care where there is a large degree of

uncertainty. Nevertheless, repeated practice with cases such as these can help students to develop strategies and complex mental representations of similar problems, which will benefit their non-analytical thinking, and make it easier for them to make quicker clinical decisions in the future.

4.4 Learning from my systematic review

My systematic review concluded that there is little evidence to show how effective online patient simulations were at improving clinical reasoning skills, as few evaluations of such tools have been published and those that were lacked methodological rigour. I found there were significant gaps in in the systematic review for tools set in primary care and resources that used reflection with feedback to facilitate clinical reasoning skills teaching, which theory suggests would be beneficial to student learning. My review suggested that newly developed online patient simulations could address clinical problems in primary care and use reflection to address these issues. Furthermore, few studies included detailed descriptions of the features of their simulations, so it was not possible to estimate effectiveness of specific possible features of simulations, such as feedback and scaffolding. Thus, my review and the empirical literature could not significantly contribute towards the logic model to explain how online patient simulations work. However, the review did suggest that there is a gap in literature to develop a logic model of how online patient simulations could improve reasoning and to conduct robust evaluations to test this theory and measure effectiveness.

4.5 Logic model for online patient simulations

Based on my learning from theory, health services research and my systematic review, I developed an online patient simulation (OPS) logic model that outlines how an online patient simulation could lead to improved clinical reasoning skills in medical students (see **Figure 4-2**). The OPS logic model not only describes how such tools could impact a student's reasoning but also how the success of that impact is directly affected by the role of the medical faculty in implementing the tools in the curriculum. My explanation of the logic model will refer to my theoretical model of students' clinical reasoning described in **Section 1.5** (**Figure 4-1**).

Figure 4-2 A diagram of the online patient simulation (OPS) logic model of how online patient simulations can improve the clinical reasoning skills of medical students

	Inputs	Programme activities	Mechanisms of change	Measurable outcomes	Impacts
Undergraduate medical students	Requirements for intervention: • Willingness to learn clinical reasoning skills using OPS. Targets of the intervention: ❖ Improving clinical reasoning skills by developing complex mental representation of illnesses (**A**).	Explicit teaching on clinical reasoning skills using educational techniques: • Multiple patient cases; • Scaffolding; • Reflection; • Feedback (**B**). Cases reflect areas of clinical importance: • Conditions that are hard to detect early in primary care/ often missed e.g. lung cancer, heart failure.	• Developing mental representations of illnesses (**B**). • Developing non-analytical reasoning skills through exposure to multiple cases (**B**). • Developing analytical reasoning skills and mitigating influence of cognitive biases through prompts for reflection during case, after case and feedback (**B**).	Data gathering skills improve e.g. • Taking a relevant and focused history. • Gathering necessary information for an informed diagnosis (**C**). Diagnostic hypotheses improve: • Adapt diagnoses to new information • Prioritise diagnoses (**D**).	• Become better doctors? • More timely and accurate diagnosis and management of patients in primary care – improving patient safety and outcomes (**D**).
Medical school faculty	Requirements for intervention: • Willingness and capacity to deliver explicit clinical reasoning skills teaching using OPS.	• Deliver OPS to students via curriculum.	• If see benefits to students' reasoning faculty may be more willing to use OPS to teach clinical reasoning skills.	• OPS implemented in curriculum and continually developed.	• Improve explicit teaching of clinical reasoning skills in medical education.

Drivers for programme
• Policy: To improve patient safety and early diagnosis of cancer, particularly in primary care
• Practice: To address errors caused by clinical reasoning.
• Education: To provide explicit clinical reasoning teaching in medical schools. To provide experiences that students may have limited access to.

Key: Letters A, B, C & D refer to corresponding letters in Theoretical model (Figure 4-1). A: Student attributes. B: Thought processes. C: Behaviours. D: Consultation outcomes.

4.5.1 Contextual drivers for online patient simulations and inputs

In **Chapter 1** I described how there are national policy initiatives in the UK and internationally to improve patient safety and reduce missed diagnostic opportunities in primary care, particularly for conditions like cancer (DoH, 2015; WHO, 2017). Evidence from clinical practice has shown that it is errors in clinical reasoning that are the main contributor to missed diagnostic opportunities (Graber et al., 2005). Providing training that explicitly teaches clinical reasoning skills in undergraduate medical education has been identified as one way to address these issues (Higgs et al., 2008). An online patient simulation tool could be used to deliver such training, as it is supported by theories of learning and clinical reasoning. My theoretical model indicated that an intervention should aim to target medical students' mental representations of illnesses, to help them improve their reasoning and have potentially fewer missed diagnostic opportunities (see Component A in **Figure 4-1** and Inputs **Figure 4-2**). Willingness from undergraduate medical students and medical faculty to respectively use and deliver explicit training on clinical reasoning skills using innovative teaching methods is also needed to create a suitable environment to develop and test an online patient simulation intervention.

4.5.2 Programme activities

To be effective online patient simulations need to include educational techniques known to improve learning outcomes, such as repeated practice with multiple patient cases, reflection, scaffolding and feedback (Croskerry, 2003; Mann, 2011). Cognitive theories and theories of learning suggest that use

of these features could improve reasoning by helping the student to reconstruct their mental representations of illnesses, allowing them to develop analytical and non-analytical thinking skills and mitigating some of the negative effects of cognitive biases on reasoning (Croskerry, 2002; Mamede et al., 2007; Merriam et al., 2012; Norman et al., 2016). Additionally, online patient simulations can be adapted to teach reasoning skills in contexts identified as being clinically important. Empirical evidence has suggested that one clinical important area is when patients present with non-specific symptoms that could be indicative of serious conditions like lung cancer (Sheringham et al., 2016; Singh et al., 2013). For online patient simulations to be used by medical students as part of their training, medical faculty would need to integrate them into their curriculum (Programme activities in **Figure 4-2)**.

4.5.3 Mechanisms of change

Online patient simulations could help students to learn clinical reasoning skills by improving their mental representations of illnesses (Component B in **Figure 4-1** and Mechanisms of change in **Figure 4-2**). They can do this through facilitating and developing students' analytical and non-analytical thinking skills. Theories of cognition and learning suggest that facilitating analytical clinical reasoning in complex clinical cases could reduce the negative influence of cognitive biases. Three cognitive biases (confirmation bias, anchoring and the unpacking principle) have been identified to be present in the decision-making of clinicians and students in previous studies and could be particular targets of online patient simulations (Kostopoulou et al., 2012; Sheringham et al., 2016). The logic model shows that an intervention that prompts students to reflect on

their reasoning, throughout a patient consultation, and provides scaffolding and feedback could help to facilitate analytical clinical reasoning (Croskerry et al., 2013). By reflecting, students would be more likely to attend to evidence that is inconsistent with their hypothesis and consider alternative hypotheses, thereby reducing the chance of confirmation bias and anchoring (Mamede et al., 2007). Reflection also encourages students to explore their hypotheses thoroughly, ensuring that they elicit all relevant information from patients, reducing the effect of the unpacking principle (Mamede et al., 2007). Scaffolding and feedback can also help students to reflect by highlighting errors in their thinking and guide them to suitable ways of approaching a case (Croskerry, 2003; Mann, 2011). Having multiple cases for students to practise with will also help students to develop key skills of non-analytical reasoning, such as pattern recognition and developing complex mental representations of illnesses. If medical faculty implement online patient simulations in their curriculum and see the benefits to students' skills, they may be willing to enrich their teaching using these methods in the future.

4.5.4 Measurable outcomes

If online patient simulations are used and there is improvement in clinical reasoning skills this should be observable in the associated behaviours and consultation outcomes around clinical reasoning, which are described in Component C and D of the theoretical model in **Figure 4-1** and the measurable outcomes in **Figure 4-2**. A student's data gathering skills should improve; in particular, their ability to take a relevant history that is focused on their diagnostic hypotheses and their ability to recognise the essential information

needed to make an informed diagnosis (Component C in **Figure 4-1** and the Measurable Outcomes in **Figure 4-2**). Additionally, a student's ability to make an appropriate diagnostic hypothesis should also improve. This could be demonstrated in their willingness to change and adapt their diagnoses based on new information and their ability to prioritise their diagnoses (Component D in **Figure 4-1** and the Measurable Outcomes in **Figure 4-2**). If medical faculty observe online patient simulations to improve a student's data gathering and diagnoses, then they may implement online patient simulations into the curriculum and develop them further.

4.5.5 Impacts of the programme

I include a section on long-term impacts of the programme that I do not propose to measure but give an indication of why I think this work is important. If behaviours and outcomes associated with 'good' clinical reasoning skills are demonstrated, they can take these skills forward in their careers and potentially become better doctors (Component D in **Figure 4-1,** Impacts in **Figure 4-2**). Clearly, there are also many other skills that a student needs to obtain to become a better doctor, such as communication skills. The specific impact of online patient simulations on whether students become better doctors may not be measurable because of other factors, such as their other teaching, which will influence their skills as a future doctor. Ultimately, it is hoped that by improving clinical reasoning that students will be able to make more timely and accurate diagnoses once qualified, which would significantly reduce patient harm and improve patient outcomes. Furthermore, it is hoped that if online patient simulations prove to be effective then they could potentially improve the

teaching of clinical reasoning skills in medical education with minimal impact on teaching capacity.

4.5.6 Testing this theory

In order to test the OPS logic model an online patient simulation intervention needed to be developed and empirically tested. The logic model describes what features such a tool should ideally include and what behaviours could be observed in an evaluation to show the tool worked. My systematic review showed that few methodologically rigorous studies have evaluated online patient simulations' impact on clinical reasoning skills for students or developed a logic model. Thus, developing such a tool following the OPS logic model and evaluating its feasibility would significantly contribute to the field.

4.6 Developing the tool

4.6.1 The context in which the tool was developed

The online patient simulation, which I contributed to the development of as part of my PhD, was not only influenced by the OPS logic model outlined in **Section 4.5** but also by external factors. These included funding of a previous related research project, feedback from the steering committee and some initial tests with medical students using a prototype tool initiated by my supervisors JS and APK, before my PhD began.

4.6.1.1 Previous funded research on the topic

The development of the online patient simulation received funding from the PRU, based on a previous study they funded, which showed that in 42% of online patient vignettes, GPs failed to elicit enough information to make an informed referral of lung cancer. Therefore, GPs' clinical reasoning skills were found to be influenced by cognitive biases (Sheringham et al., 2016). Funding was provided for my PhD to develop a patient vignette based training tool to support clinical reasoning skills and evaluate this tool with a trial. The previous project also provided a set of patient-actor videos that had been developed with patients and clinicians, which reduced the burden of development time for new patient cases.

4.6.1.2 Initial feedback from steering committee

To explore the idea of developing an online patient simulation training tool to improve clinical reasoning skills, my supervisors and I sought feedback from a steering committee for the wider project that included a panel of experts and a lay member (see **Appendix 2** for members). They suggested that the tool should address specific cognitive biases and should have a variety of complex clinical cases. They suggested the clinical cases should include lung cancer, as this diagnosis is prone to missed diagnostic opportunities. However, they felt it would provide more value to students and faculty as a learning tool if it presented a range of conditions. They suggested it focused on patients with respiratory and related symptoms, indicative of a variety of serious conditions, such as lung cancer, heart failure and chronic obstructive pulmonary disease

(COPD). All these conditions can present with vague non-specific symptoms and could be subject to missed opportunities for diagnosis in general practice. The committee also suggested the tool should firstly be targeted towards medical students, who may have more time for learning and may be more open to learning these skills via this novel method than GPs. Indeed, their suggestions fit with those in the wider literature, discussed in **Chapter 1.** General practice is a context in which clinicians and students are faced with complex diagnostic challenges, such as negotiating diagnostic hypotheses frequently and rapidly, under time and resource constraints (Cheraghi-Sohi et al., 2018; Rubin et al., 2015). Furthermore, students have limited practice using their reasoning in general practice settings. Therefore, there is potentially a need for complementary reasoning teaching for students, using non-face-to-face learning methodologies (Harding et al., 2015). Enabling students to gain a better understanding of how to use clinical reasoning skills in general practice and what the challenges can be, may help them to form better skills earlier in their training and avoid the pitfalls of becoming fixed in their thinking (Audétat et al., 2012).

4.6.1.3 Initial tests using a prototype tool with medical students

As a pilot phase, I undertook a small study analysing data from fifth year medical students who had used a prototype online medical education training tool developed by my supervisors JS and APK before my PhD. The tool used the videos of actors representing patients that were developed by the previous research project described in **Section 4.6.1.1**. Students were required to formulate a differential diagnosis and management plan for patients presenting

with respiratory symptoms to their GP. It was not a simulation, as it did not require the student to gather information from a patient, but it required students to use their clinical reasoning skills to reach a diagnosis, so it was considered informative for the development of an online patient simulation tool to improve reasoning. The prototype was used by an urban medical school as part of their online learning. Case of the Month provides students with patient cases related to their curriculum. I analysed fifth year medical students' success at formulating a diagnosis and their feedback on the case. I analysed data collected for the case from January-April 2016; 41 students out of 67 completed the case and 32 gave written feedback.

Overall, I found that students struggled to formulate a differential diagnosis and management plan for patients. However, four students commented that one of the benefits of the case was that it helped them to practise diagnosing and managing patients in this context. Furthermore, 27 students commented that they felt it helped them to identify gaps in their knowledge, such as the NICE guidelines for urgent cancer referral. This suggested that the tool could provide students with more practice formulating a differential diagnosis and managing patients in this context, to improve their decision-making and knowledge. The main problem with delivering teaching in this way was technical difficulties, such as the videos not working. Therefore, the key recommendations from feedback and this analysis were that the tool should include a variety of respiratory cases that require students to know and learn key clinical guidelines. It also emphasised the importance of ensuring the technology works appropriately to retain participants.

4.7 Description of the tool

Based on the logic model, the external factors to my PhD and feedback from the prototype, outlined in **Section 4.6**, I contributed towards the development of an online patient simulation learning tool. My particular responsibility was to share evidence from empirical studies and theories about design features and to lead the evaluation components. Others developed the content and functionality. As part of the team, I played a role in most decisions about the design and development of the tool. The tool was named electronic clinical reasoning educational simulation tool (eCREST) by the research team. It was designed specifically for this project and addresses how to manage patients presenting to their GP with non-specific respiratory symptoms. In the online patient simulation, students are presented with videos of three actors, representing patients, to simulate a doctor-patient consultation. The patients are presented via video to make the experience feel realistic, as this is likely to improve engagement with the tool (Kolb, 1984). The videos were sourced from JS, who used these 'patient' videos in a previous study (Sheringham et al., 2016). The role of the student is to act as a junior doctor and gather information from the patient, while continually reviewing their diagnosis. At the end of each case students are asked to formulate a differential diagnosis and management plan. The patient cases are designed to be typical of cases seen in general practice, in which symptoms are vague and the diagnosis is unclear. The first case presents a patient where lung cancer was an important diagnosis, the second heart failure and the third pneumonia. The clinical content of the patient cases, including their symptom profile and differential diagnosis is outlined in

Appendix 3. **Appendix 3** also contains the login details for access to the administrator account of eCREST, where the tool can be viewed.

Figure 4-3 provides an overview of the user process of the intervention. Students begin by selecting a patient and can select any patient in any order. Students watch a short video of a patient describing their problem (**Figure 4-4**). Students then rate how concerned they are about the patient and choose their top five diagnoses from a predefined list (**Figure 4-5**). The students gather data from the patient by selecting questions from a list, to which there is a video response from the patient. Questions are based on the Calgary-Cambridge guide to the medical interview, which is taught at many medical schools (Kurtz, Silverman, Benson, & Draper, 2003). There is no limit to the amount of questions students can ask and there are approximately 30-40 questions in each case (**Figure 4-6**). They can ask up to six questions at a time and are then prompted to review their diagnosis. They can change their differential diagnosis by adding, removing or re-ordering their diagnoses; they must explain why they chose to change their diagnoses (**Figure 4-7**). They can also select to receive up to eight results from a range of physical examinations and bedside tests, in text format. At the end of each case students are asked to list their final diagnoses and explain why their choices changed (or not) throughout the consultation. They choose how to manage their patient by selecting further tests and follow-up options from a list. Students then receive video feedback from a GP and GP registrar and are provided with links to relevant guidelines (**Figure 4-8**). They also receive a PDF showing what they did in the consultation, links to relevant guidelines and a clinical reasoning toolkit from the Society to

Improve Diagnosis in Medicine (The Society to Improve Diagnosis in Medicine, 2019). After case feedback they are directed back to the waiting room to complete the other patient cases. After completing all three cases, students are asked to reflect on what they have learned from eCREST. Each case takes approximately 20-30 minutes to complete.

Figure 4-3 A diagram of the user process for eCREST

1. Select patient case from waiting room	2. Watch video of the patient's presenting complaint and rank top 5 diagnoses	3. Select up to 6 questions to ask patient from a predefined list
4. Review top 5 diagnoses; justify choice. Steps 3 and 4 can be repeated until satisfied with information	5. Select results of up to 8 physical examinations or bedside tests. Review top 5 diagnoses; justify choice	6. Ask a final 2 questions or receive 2 more results from the physical examinations or bedside tests
7. Select top 5 diagnoses and justify choice; select further tests and follow-up for the patient	8. Receive video feedback from GP and GP registrar on final top 5 diagnoses, further tests and follow-up	If completed 1st or 2nd case return to step 1 and select next patient case
		If completed 3rd case reflect on performance and clinical reasoning

Figure 4-4 Screenshot of the virtual patient cases in eCREST
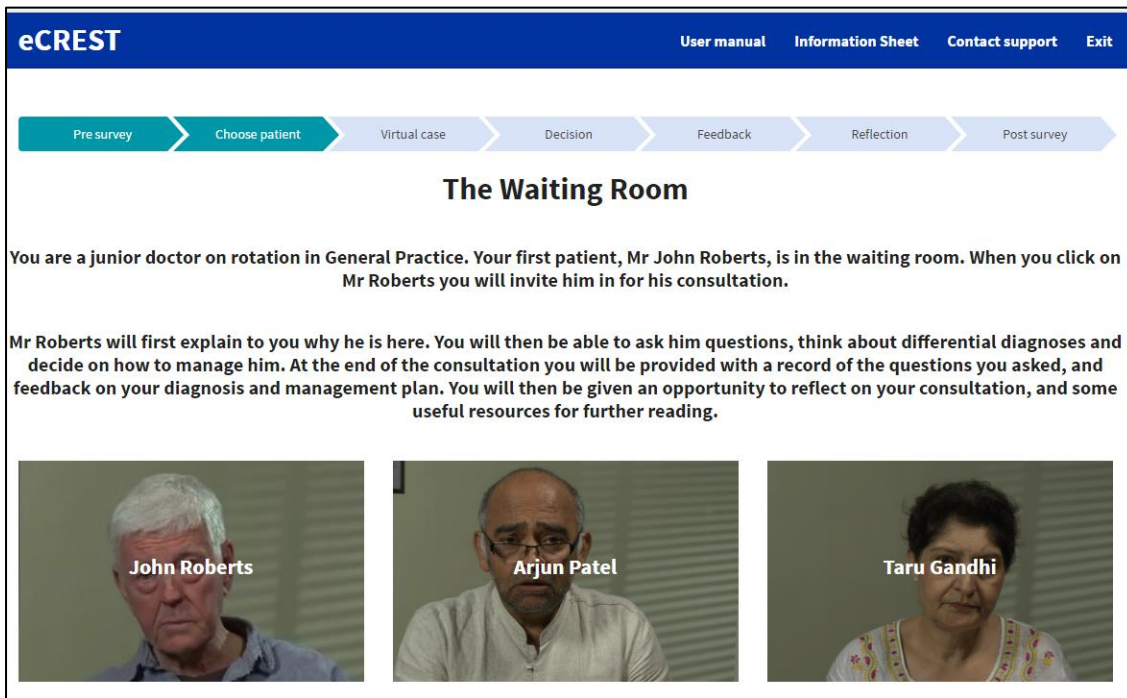
Figure 4-5 Screenshot of how students select their initial diagnosis
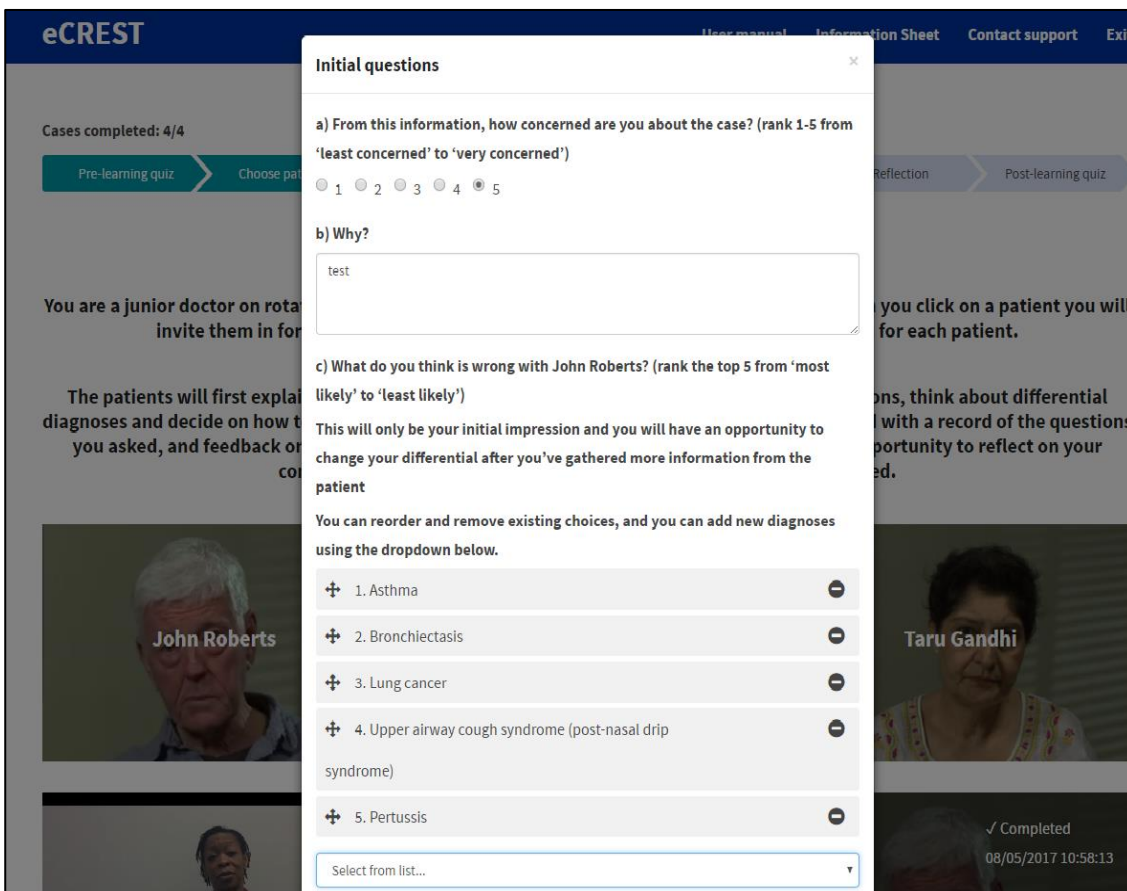
Figure 4-6 Screenshot of how students can ask the patients questions in eCREST
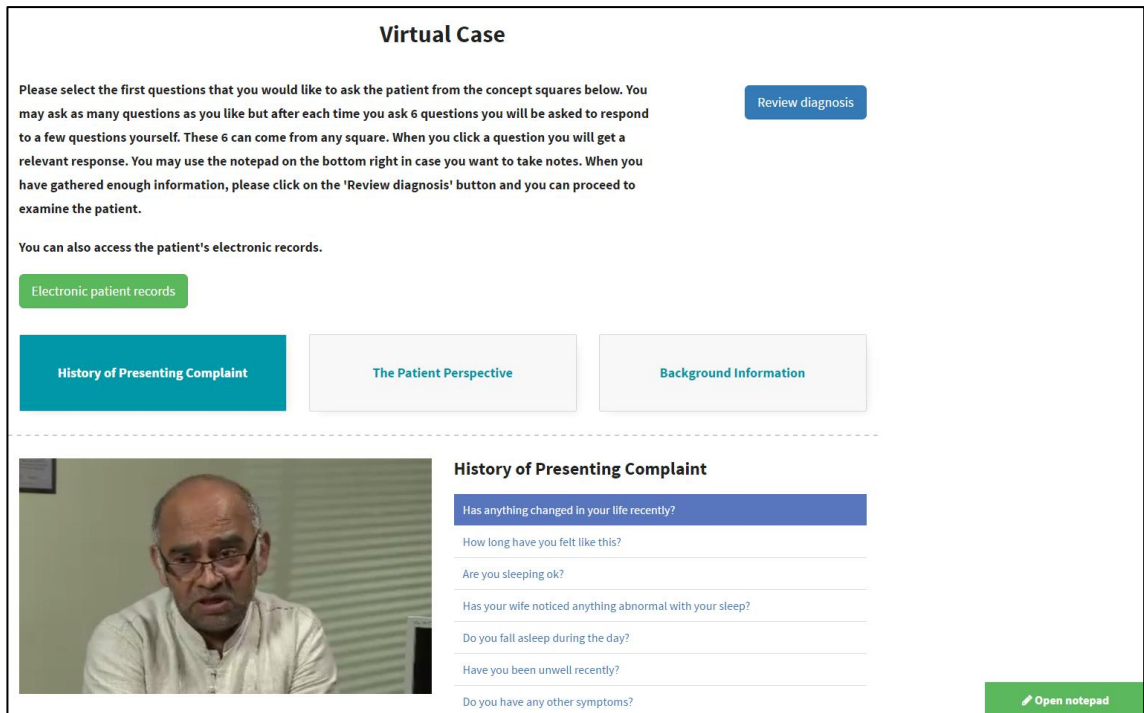
Figure 4-7 Screenshot of how students are prompted by eCREST to review their diagnosis
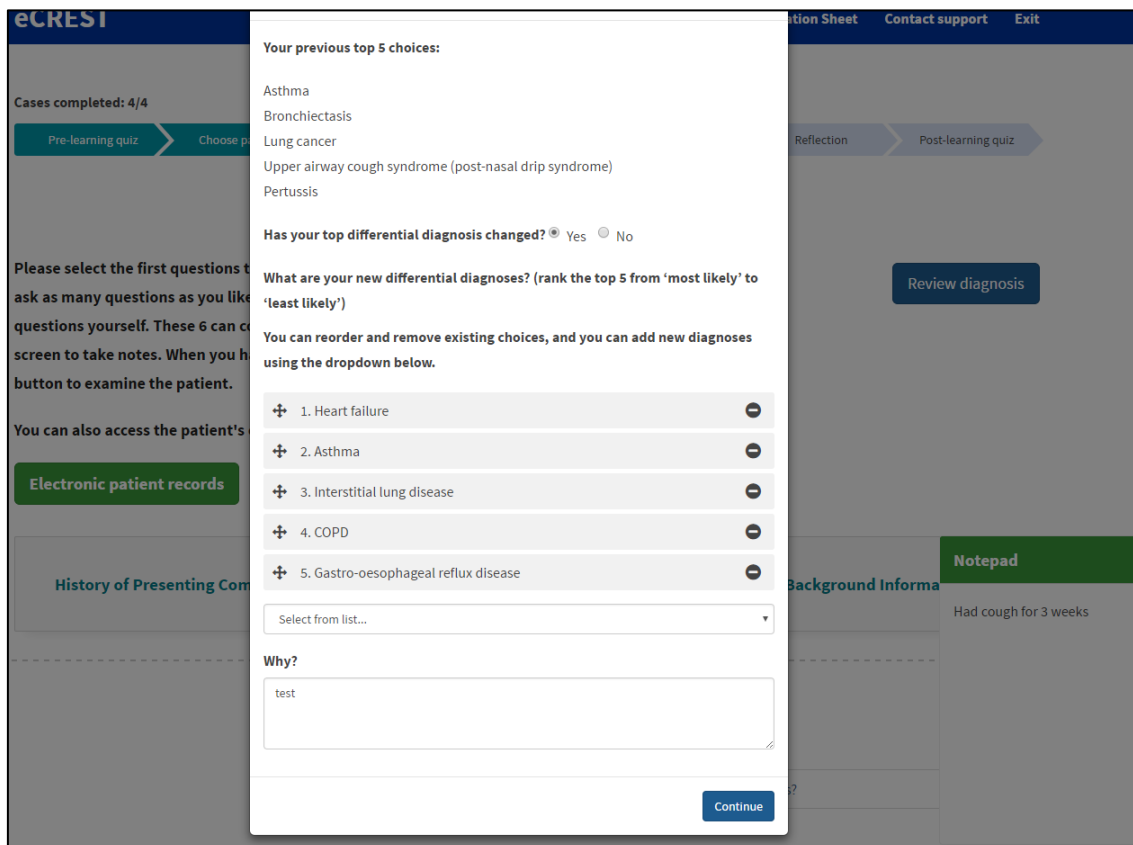
Figure 4-8 Screenshot of how feedback is provided to students in eCREST
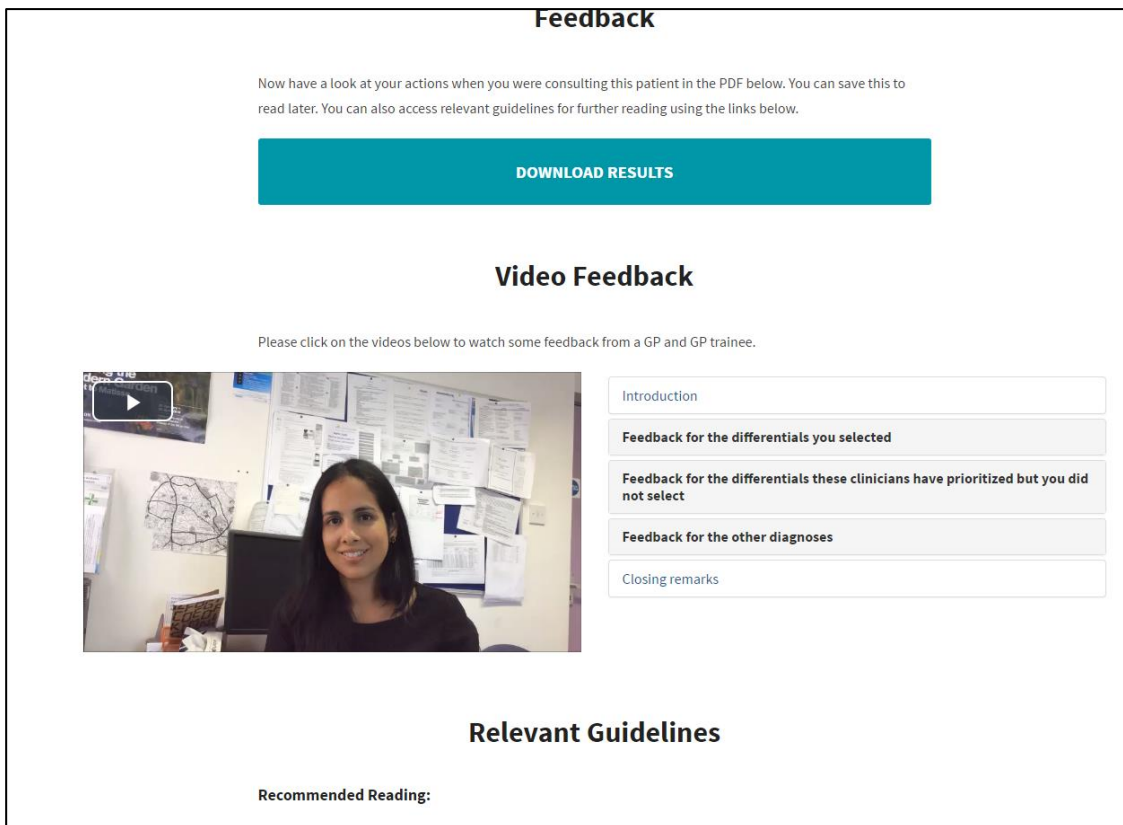
4.7.1 Explanation of why these features were included in eCREST

eCREST was designed this way to facilitate reflection and provide scaffolding and feedback, which the OPS logic model (**Section 4.5**) suggested would be needed to improve students' clinical reasoning skills. Reflection is facilitated and scaffolding provided, throughout the virtual consultation by asking students to rank their top five differential diagnoses and justify their choices. This follows the process described by the iterative hypotheses testing approach to teaching reasoning, in which diagnoses are continually refined and requests for patient information are justified (Kassier, 1983). Students are also asked to reflect on what they have learned at the end of eCREST. Prompting students to re-evaluate their diagnoses and asking them why they are choosing

their diagnoses, forces the students to reflect on why they are making these decisions, think of their differentials early in the case and consider other options. This also encourages reflection-in-action, which is recommended by Schön as useful for learning skills (Schön, 1983). Feedback on student performance is facilitated by video feedback from a GP and GP registrar on what they should have chosen for their final differential diagnosis and management plan of that patient. Providing this feedback can help students to reflect-on-action and evaluate their performance to consider what they might improve on for future cases; this process can help students to restructure their mental representations of illnesses (Mamede et al., 2008; Rutherford-Hemming, 2012; Schön, 1983).

4.8 Patient and Public Involvement (PPI) in the project

Patients were involved with the design of the patient vignettes used in my study, as a small group of patients co-wrote the scripts of the vignettes and helped to identify pertinent clinical and behavioural characteristics for the simulated cases. This was undertaken before my PhD began for the previous research project led by my supervisor JS (Sheringham et al., 2016). Patients and the public on the Research Advisory Panel (RAP) facilitated by the NIHR CLAHRC North Thames at UCL were also consulted regarding my PhD. This panel consisted of around 10 lay members from a diverse range of backgrounds with expertise in a number of different areas. Before my PhD began, my supervisor (JS) presented the prototype of the tool and the research idea to the panel. They provided feedback on the rationale for the research and acknowledged the importance of addressing diagnostic errors in primary care.

The RAP also had feedback on the content of the patient cases. They wanted to ensure that the patients presented were complex and covered patients from a range of ethnic backgrounds, which was taken into consideration in shaping and selecting the final patient cases for the tool. They expressed concern around how the tool will feedback the 'right answer' regarding diagnosis to students. To address this issue the feedback reflected the preferred approaches of the GPs who gave the feedback, but with explicit acknowledgement that this could vary depending on the clinician.

I returned to the RAP meeting in the middle stages of my PhD and presented the patient cases and plan to evaluate the research tool. There was feedback from the panel that they recognised the rationale of the study and the need to support future clinicians in their clinical reasoning to try and minimise the possible negative effects of missed diagnostic opportunities. For example, they talked about how they felt GPs can be tempted to jump to a diagnosis and not listen for new information. However, they also questioned whether it was possible to change these ingrained thought processes. The panel also had feedback regarding the patient cases. They would have preferred the cases to be real patients, who brought their own story and experience to the cases, to help medical students become aware of the different experiences of patients. One of the panel members commented that they felt it was important that students are "[Reminded that] they are treating people not symptoms". One suggestion to address this issue was to add some patient feedback on what they as a patient would have expected and wanted from the consultation, in the feedback to the student. I thought this was an excellent suggestion, as it could help to remind

students of the importance of reflecting on how their reasoning is affected by the patient's story and attributes of the patient. Without the input from the panel I would not have thought about adding this potentially very helpful element to the feedback. However, due to the timing of the consultation with the panel it was not possible to add this feature to the current patient cases. On reflection, I would have liked to meet with patients in the earlier stages of my PhD, when the tool was being developed, rather than after development, so that their experiences could have further informed the tool. The panel's suggestions to use real patients were also not taken forward. This was partly because of ethical concerns but mainly because simulated patients are designed to represent specific clinical, social and behavioural characteristics. Actors are trained to adapt their delivery, manner and appearance to portray these. The wider research team did respond to the panel's suggestions, providing reasons why they were or were not taken forward (via the chair). I think that this communication and acknowledgment of their feedback was helpful to maintain a good working relationship with the panel for future research. Indeed, the panel's feedback will inform potential future iterations of the tool.

4.9 Piloting eCREST

A co-creation approach was taken to developing eCREST. The prototype version of eCREST (described in **Section 4.6.1.3**) was co-developed by peer assisted learning scheme (PALs) students with patient consultation. The content of all the patient cases were developed by GP registrars in collaboration with a wider group of expert clinicians (Asarbakhsh & Sandars, 2013; Sanders

& Stappers, 2008). I also shared with this group of GP registrars, iterations of the OPS logic model and proposed evaluation tools (described in **Chapter 5**) for feedback. Clinicians and students also tested the first fully developed version of eCREST to determine the usability and appropriateness of the tool for medical students. Initial feedback was received from five experts who rated the tool's design/layout, content and ease of navigation on a scale of 1-10, with 10 being the highest. The lowest score was the tool's ease of navigation ($M = 7.6$). The design/layout and content scored higher ($M = 8.4$ and $M = 8.5$ respectively). A small sample of medical students, who were in their fourth ($n = 2$) and sixth year ($n = 4$) of study in an urban medical, also rated eCREST using the same scale. Similarly to experts, students rated ease of navigation as the poorest element of eCREST ($M = 6.3$). They scored the design/layout highly ($M = 8.5$) but the content was rated lower by students than experts ($M = 7.8$). Feedback from the students and experts indicated that clearer instructions on how to proceed through the cases were needed. It was suggested that the layout of the questions could be improved by using drop down menus and that 'take home messages' could be added in the feedback. These features were all added to the tool. Another key recommendation, was that eCREST was more suitable for final year undergraduate medical students, due to the complexity of the cases; fourth year students struggled with the clinical content more than the final year students.

4.10 Challenges to developing eCREST

There were practical challenges to developing eCREST, such as the length of time it took to develop and the inputs needed from a large team of

people with different expertise. I found it took a considerable amount of time (approximately a year and a half) to develop four patient cases. I also found that it required the involvement of a large network of people to create the cases, and to develop and test the tool. The creation of the clinical cases involved the work of several GPs, GP registrars, medical educators and other clinicians, to ensure the cases were appropriate and valid. Furthermore, it required finding a web developer and liaising continually with them to transfer the content to the website and manage the technical aspects. It also required recruiting several students to test and provide feedback to ensure the online patient simulation was user-friendly. This means that medical schools seeking to develop their own tools could struggle to invest the time and resources needed to develop digital innovations in teaching. However, the learning from this PhD, including the description of how eCREST was developed and logic model, could help future developers to streamline the development process. Furthermore, eCREST itself could be used by medical schools to reduce time spent on development.

A further challenge to the development of the tool and the PhD more generally was that the PhD was produced as part of a wider project. This was helpful for the project, as it meant that the patient vignettes had mostly been developed before the PhD began and that there were already relationships built with key stakeholders in the PhD, such as the steering committee members and medical educators across various medical schools. However, it also limited some aspects of the PhD. For example, because the patient vignettes had already been largely developed, eCREST had to be designed to

fit with these vignettes, and this may have limited some of the clinical scenarios that the GP registrars tried to create when forming the eCREST patient cases. Undertaking PhD research within larger projects is now relatively commonplace; I found the best way to acknowledge the challenges to this approach, is to be explicit in describing the decisions you make and those made by others and to provide the rationale for both.

4.11 Summary

This chapter integrated evidence from theory (outlined in **Chapter 1**) and empirical evidence (outlined in **Chapters 1 & 3**) to develop an OPS logic model that explains how an online patient simulation training tool could improve the clinical reasoning skills of medical students. I described how this theory and external factors to my PhD influenced the development of eCREST. However, in order to test the OPS logic model and estimate the feasibility of online patient simulations, an evaluation of eCREST is needed. My systematic review (**Chapter 3**) indicated that a robust evaluation of an online patient simulation learning tool would significantly add to the literature given that there is a lack of high methodological quality evaluations in the current literature.

5 Methods: Feasibility RCT

I described in **Chapter 4** how eCREST was developed based on the OPS logic model. In this chapter, I present how I addressed research question three of my thesis, which was to estimate the feasibility, acceptability and potential impacts of eCREST on clinical reasoning skills via a feasibility RCT.

5.1 Background

In **Chapter 3**, my systematic review found that relatively few RCTs of high methodological quality have measured the impact of online patient simulations on undergraduate medical students' clinical reasoning skills. As RCTs are the gold-standard method of evaluating interventions, further studies that use this design could help to build a more methodologically robust evidence base, to estimate the effectiveness of online patient simulations (Campbell et al., 2000). However, it can be difficult to establish cause and effect and make generalisable conclusions using RCT designs in medical education because of the complexity of the interventions and context in which they are set. A significant challenge is that medical education interventions are context-dependent, with outcomes varying depending on how they are delivered, who they are delivered by and the attributes of the learners. Moreover, educational interventions consist of several components that interact with each other (Sullivan, 2011; Wong, Greenhalgh, Westhorp, & Pawson, 2012). For example, in eCREST the learning objectives, the content of clinical cases, the interactivity of the cases, the use of videos to present patients, the reflection on objectives

and skills, the feedback and the pre and post outcome assessments, all contribute to student learning. Furthermore, it may be unethical to assign students to a control or placebo condition, as it could have an effect on their exams (Sullivan, 2011). Indeed, ethical concerns tended to be the main concern for medical schools that were approached to take part in my study, as ensuring the courses were equitable was a priority. Thus, it is challenging to use an RCT design to understand how and why and for who eCREST would be most effective. That is not to say that exploring effectiveness using RCTs in medical education is redundant. It is important for the complexity of the setting, intervention and participants to be captured, to help researchers theorise how and why such interventions may or may not have an effect (Wong et al., 2012). A useful first step in the approach to evaluating eCREST in such a complex setting would be to conduct a feasibility RCT. A feasibility RCT would capture wider data on context, recruitment, participants and indicate whether a full RCT would be appropriate given the context. Furthermore, the use of qualitative research methods to explore the mechanisms involved in learning could help to provide a richer understanding of how and why eCREST is effective (See **Chapter 7**).

A feasibility RCT would also be useful to use as an evaluation design for eCREST because the current evidence base for effectiveness of such interventions is limited. It was not clear from my review: what the expected uptake and dropout rates would be if eCREST were to be used by students (particularly if it was not integrated into the curriculum); whether students would accept eCREST and how to assess clinical reasoning skills using valid and

reliable measures. Therefore, an evaluation to assess the feasibility of using an RCT to estimate the effectiveness of online patient simulations on clinical reasoning skills was needed before a full-scale definitive trial could be used to measure effectiveness. Indeed, a feasibility RCT would help to provide evidence of proof of concept for eCREST, which is likely to be required before it could be implemented into curricula or evaluated in a full-scale trial.

5.2 Aims and objectives of the feasibility RCT

One of the aims of my thesis was to evaluate a training tool based on evidence and theory based to improve the clinical reasoning skills of medical students. A further aim was to develop novel ways of assessing reasoning skills. To achieve these aims I needed to answer my third research question, which was to assess the feasibility, acceptability and potential impacts of an online patient simulation and develop, and estimate the validity and reliability of my clinical reasoning measures. Therefore, the objectives of this study were to:

a) conduct a feasibility RCT study to determine the feasibility and acceptability of conducting an RCT study to evaluate the impact of eCREST on medical students' clinical reasoning skills, compared to usual teaching;

b) develop and validate measures to assess clinical reasoning skills;

c) analyse the potential impacts of eCREST on clinical reasoning skills using these measures by comparing the outcomes from the intervention and control group.

5.3 Study design

A feasibility RCT design was used based on the guidelines from The Medical Research Council on the evaluation of complex interventions, and from NIHR (Campbell et al., 2000; Craig et al., 2008; NIHR, 2016). I used these guidelines because eCREST is a complex intervention, as it was intended to be delivered in a complex setting where several other teaching interventions were being undertaken at the same time (Campbell et al., 2000). I chose to do a feasibility trial rather than a definitive full-scale RCT (phase III trial) as it was necessary to understand the components of the trial, such as recruitment, sample size, characteristics of the outcome measures and how the trial could be implemented in practice. This would help to determine whether it would be possible to conduct a definitive full-scale RCT, which could assess the effectiveness of eCREST on clinical reasoning skills (Campbell et al., 2000).

Feasibility trials are distinct from pilot studies, which are smaller scale versions of a definitive full-scale trial. My study was a feasibility trial as it determined:

- The number of eligible medical students available

- Willingness of medical schools to recruit participants

- How students were recruited

- Willingness of medical students to be take part (uptake)

- Follow-up rates (retention)

- Completion rates of eCREST

- Duration of intervention

- Acceptability of eCREST to medical students

- The standard deviation of an outcome measure that can be used to assess clinical reasoning skills, to undertake a power calculation to estimate sample size,

- Characteristics of the outcome measures (Arain, Campbell, Cooper, & Lancaster, 2010).

5.3.1 Setting

Data collection ran from 1st March 2017- 28th February 2018. The study took place in three medical schools in universities across the UK: School A (SA); School B (SB) and School C (SC). The medical schools in which the study took place varied slightly. SA and SB implement an integrated/systems-based curriculum, in which they take a bodily system, such as the circulatory system, and consider the anatomy, physiology, biochemistry, pharmacology and pathology of that system together. The first 1-3 years are generally classroom based (lectures) and clinical placements are more common the latter years of study. These medical schools were also based in urban areas. SC was a rural based medical school and followed a problem-based learning (PBL) curriculum, in which students see patients from the first year of training and are given medical cases to resolve through group work. All universities used the Calgary Cambridge Guide (Kurtz et al., 2003) as their consultation skills model. All sites used online learning tools in their clinical training, but none have used interactive patient simulation cases set in primary care, in which the sole purpose is teach clinical reasoning skills, rather than any other type of clinical

or procedural skill or knowledge. More information about the different medical schools can be seen in **Table 5-1**.

Table 5-1 Study settings

Medical School	Setting	Course length (years)	Approach to teaching	Primary care placement in final year	Online learning platform used by school
SA	Urban	6	Integrated/systems based	4 weeks placement.	Moodle
SB	Urban	5	Integrated/systems based	6 weeks placement.	Moodle
SC	Rural	5	Problem-based Learning	One day a week in primary care.	Blackboard

5.3.2 Participants

All undergraduate medical students in their final year of clinical training at SB, SA and SC medical schools were eligible to take part in the study. Final year medical students were chosen based on consultation with the medical school education leads, who felt that the level of difficulty of the cases was more suited to those in their final year, and feedback from the testing of the first version of eCREST with PALs students (**Section 4.8**). I confirmed with the medical school leads of each university that final year students had the prerequisite knowledge to manage and refer patients presenting with respiratory and related symptoms in general practice. Some characteristics of the participants varied due to timetabling differences between universities and because students were recruited over two academic years; see the recruitment

Section 5.4.3 for further details. See **Table 5-2** for details about how clinical reasoning was taught at each medical school and the different placements students were on during the eCREST study, across the two academic years the study was carried out.

Table 5-2 Clinical reasoning and other teaching occurring during eCREST at each medical school

Medical school	How was clinical reasoning taught?[1]	Placement Academic year 2016/2017	Placement Academic year 2017/2018
SA	Clinical teachers provide feedback on performance and progress of students through Clinical Reasoning Discussions. Lectures on clinical reasoning in year 5. Reflective practice essays.	Pre-elective/ student selected component.	GP placement.
SB	PBL, primary care placement (interviewing and examining patients) and during clinical skills teaching which is integrated into each module.	Preparation for practice/ student assistantship.	Student selected component/ anaesthesia & critical care/ inpatient medicine & acute and emergency medicine.
SC	PBL - GP tutor recruits patients for them to examine each week and provide 'hot cases' where student is first to assess patient.	n/a – did not take part.	Emergency care/ student assistantship.

[1]Taken from the medical schools' websites and student handbooks.

5.4 Procedure

5.4.1 Intervention

The intervention group completed three patient cases online and received feedback. For a full description of the patient cases and eCREST see **Appendix 3** and **Section 4.7**.

5.4.2 Comparison group

The control group were assigned to receive teaching as usual, so did not receive any additional teaching from eCREST during the study or any additional teaching on reasoning. The control group were informed that they would have access to eCREST at the end of the study period.

5.4.3 Recruitment

Initially, my supervisors (JS and APK) and I approached each medical schools to discuss the possibility of fully integrating eCREST and the RCT design into the curriculum, thereby making participation compulsory. However, this approach was not taken because the medical schools had concerns about unfairly disadvantaging students who would not receive any extra training in the control condition and because eCREST was an untested tool. It was agreed that students would be recruited to take part on a voluntary basis at all medical schools and eCREST would be extracurricular. Students were recruited via email invitation from the medical school, newsletters, verbal invitation on their teaching days, social media and recommendation from student representatives.

Students were initially recruited from two medical schools, SA and SB. Initial recruitment took place in the spring and summer terms of the 2016/2017 academic year, and students who took part in this stage were defined as being in cohort 1. Due to low uptake, more students were recruited the following academic year, and defined as being in cohort 2. An additional medical school, SC, was also added to the study in the next academic year. I added this group to the study because they follow a different type of curriculum to SA and SB, so some comparisons between types of curriculum could be made. Furthermore, it could have increased the sample size and SC's head of department of medical education had expressed interest in participating in the feasibility RCT.

The second cohort of students were recruited from all three medical schools, in the autumn and winter terms of 2017/2018 academic year. Cohort 2 students were recruited mostly via advertisements on their eLearning platforms (such as Moodle). I worked with administrators at each medical school to place eCREST on their eLearning platforms. For SA students, eCREST was timed to occur in line with their primary care clinical placement. I have included a detailed description of how successful recruitment across the two cohorts using these different strategies was in **Section 6.1.2**. I included it in the results chapter, as this was a feasibility trial and one of its outcomes was to assess willingness of participants to be recruited. Details of the proposed reasons for the low uptake in cohort 1 and learning on how best to recruit medical students are also described as part of the feasibility results in **Chapter 6**.

5.4.4 Protocol

I present a flow diagram showing the research protocol in **Figure 5-1**. Students took part in the study online and remotely, and were able to access the patient cases and the assessments via the eCREST website. Students were provided with the link to the eCREST website (http://silverdistrict.uk/ecrest/). They were asked to read an information sheet and give their consent before registering with eCREST. Registration involved providing their name, student ID number, email address, age, gender, previous degrees, where they heard about the study and they were asked to identify the medical school that they were studying in. They needed to create a username and password for future login. Students were then asked to complete two baseline surveys: one assessing self-reported clinical reasoning skills via the Flexibility in Thinking scale (FIT); and one assessing respiratory and related medical knowledge via MCQs. Students received a score out of 12 to inform them of how many MCQs they answered correctly. The computer then randomly assigned students to the intervention or control group using simple randomisation (see **Section 5.4.5**). The intervention group received three patient cases in eCREST and were given one week to complete them. They received email reminders to complete eCREST three and seven days after registering. The control group received teaching as usual and no further intervention. Both groups were sent an email one week after registration with a link back to the eCREST website to complete the first follow-up (Time 1) assessments. They were given one week to complete the surveys (FIT survey for both groups and additional acceptability survey for the intervention group) and were sent reminders after three and seven days. Students in both groups were sent a link to the eCREST website

one month after registering and asked to complete the second follow-up assessment (Time 2), which included a new patient case in eCREST – case 4, the MCQs and the FIT survey). They had a week to complete the Time 2 assessments. The one-month duration between assessments was chosen to estimate the sustainability of any learning from eCREST over this time period. Both groups received a £20 Amazon gift card via email if they completed the follow-up assessments at Time 1, and an additional £10 voucher if they completed the follow-up assessments at Time 2.

Figure 5-1 Participant flow through the trial and assessment procedure

```
┌─────────────────────────────────────────┐
│      Participants eligible to take part  │
│                SA, SB, SC                │
└─────────────────────────────────────────┘
```

Participants eligible to take part

SA, SB, SC

Recruitment and consent

Information sheet and electronic consent form

Baseline data collected: age, gender, previous degrees, university, where they heard about the study, FIT and MCQs

Excluded

Did not volunteer

Did not consent

Randomisation

Simple randomisation to 2 groups by computer

Control group

No intervention - teaching as usual

Intervention group

eCREST (complete 3 patient cases)

Follow-up Time 1

1 week after registration. FIT survey. £20 amazon voucher.

Follow-up Time 1

1 week after registration. FIT survey and acceptability survey. £20 amazon voucher.

Follow-up Time 2

1 month after registration. FIT survey, MCQs and 1 patient case. £10 amazon voucher.

Follow-up Time 2

1 month after registration. FIT survey, MCQs and 1 patient case. £10 amazon voucher.

5.4.4.1 How the study procedure was developed

This procedure was developed based on the piloting of an initial procedure for the trial. A small sample of 10 students from SB were recruited in March 2017 to take part in the study. For these students, the follow-up at Time 1 consisted of the FIT survey, MCQs and acceptability survey, rather than just the FIT and acceptability survey. They also did not receive a score of how many MCQs they got correct. At Time 2 students in this pilot group only received patient case 4 if they were in the intervention group.

The protocol was amended to the current procedure based on my review of this initial recruitment experience and from feedback. The MCQs at Time 1 were removed due to the repetition of some of the knowledge questions. There was a concern that students may get the right answer because they had remembered the correct answer from the previous questions. I included students' scores on the MCQs based on feedback. It was felt the students would be more motivated to take part if they could test their knowledge. I also provided case 4 to both the intervention and control group at the follow-up assessment at Time 2, so that performance on a further patient case could be compared between groups. In the pilot of the procedure I did not do this because I was unclear on which aspects of their performance on case 4 to assess; the development of the key features measures (see **Section 5.7.3**) shortly after, allowed this to be possible. The 10 students recruited from SB in March 2017 were excluded from data analysis presented in **Chapter 6**, as they underwent a different procedure.

5.4.5 Randomisation

Randomisation was performed after students consented to take part, registered on the eCREST website, and completed their demographic information and baseline assessments. The allocation ratio for randomisation was 1:1, into either the intervention or the control group. The web developers undertook the randomisation. The randomisation of students in the final study was not precisely 1:1 as five students were mistakenly allocated to a separate group on the eCREST website that automatically gave them access to the three cases. Therefore, I allocated these students to the intervention group, as they completed cases that the intervention group would have been exposed to anyway. Given this was a small number of students, it was unlikely to introduce bias into the data and these students were included in analyses.

5.4.6 Blinding

I was blind to group allocation, as this was completed by the computer algorithm the web developers created. It was not possible to blind participants to their allocated treatment, as it would have been obvious to the students whether they received the patient cases or not. I analysed the data without being blinded to group allocation but I attempted to attenuate any possible bias by following my data analysis protocol outlined in **Section 5.9**.

5.5 Ethics

Ethical approval was obtained from SA Research Ethics Committee (ref: 9605/001; 31st October 2016), Institute of Health Sciences Education review

committee at SB (ref: IHSEPRC-41; 31st January 2017) and the Faculty of Medicine and Health Sciences Research Ethics Committee at SC (ref: 2016/2017 – 99; 21st October 2017).

5.6 Sample size

As this was a feasibility trial, a sample size calculation was not required (NIHR, 2016). To guide recruitment, however, I calculated how many participants would be needed to achieve a moderate effect size ($d = 0.5$) on the self-reported clinical reasoning skills measure the FIT sub-scale. Power analysis revealed that a two-sided test, with $\alpha = 0.05$ would have 80% power to detect a moderate effect with 63 people in each group, assuming a standard deviation of 10.9 (calculated from a previous study) (Lee et al., 2010). Thus, my aim was to recruit at least 168 medical students to my study, assuming 75% completion, as my systematic review indicated that almost all studies had less than 25% of students drop out in their studies.

5.7 Outcome measures

The primary outcomes of this study were the feasibility and acceptability of conducting an RCT in this context. Given that effectiveness could only be estimated in a feasibility trial, the secondary outcomes of this study were clinical reasoning skills, as well as consultation outcomes and knowledge.

5.7.1 Demographic characteristics

Demographic characteristics of participants were measured at baseline including: age, gender, medical school and previous degrees obtained.

5.7.2 Primary outcomes

5.7.2.1 Feasibility

Feasibility was measured by assessing the uptake, follow-up, recruitment method, completion and duration of intervention. **Table 5-3** describes the feasibility outcomes and how they will be measured in the data. Duration of the intervention was assessed to ensure cases could be completed in a reasonable timescale.

Table 5-3 Feasibility outcomes

Dimensions of feasibility to be measured	Outcome measure
Uptake	Numbers registered out of all eligible.
Follow-up	Numbers completing T1 and T2 out of all registered.
Recruitment method	Students' reports of how they were recruited and motivation to participate in the trial (student participant survey). Numbers of students recruited by each recruitment method out of all those recruited.
Completion	Numbers completing 3 cases as a proportion of all those in the intervention group.
Duration of intervention.	Median (and inter-quartile range) time spent on each of the 3 cases.

5.7.2.2 Acceptability of eCREST to medical students

Acceptability was measured by a survey I developed, which consisted of eight questions, and was based on previous research (Asarbakhsh & Sandars, 2013; Kleinert et al., 2015). It established students' perceptions of: the navigability of eCREST; the level of difficulty of the clinical content; their perceptions of the usefulness of eCREST for learning clinical reasoning skills; their views on whether eCREST should be integrated into their curriculum; their motivation for taking part and suggestions for the improvement of eCREST. The first 6 questions were rated on a Likert scale from 1 to 5, where 1 = strongly disagree and 5 = strongly agree. The question that asked what their motivation was for taking part was a multiple-choice question. This was followed by an open-ended question asking for their suggestions of improvement. This survey can be found in **Appendix 4**.

5.7.3   Secondary outcome measures

5.7.3.1 Clinical reasoning skills

5.7.3.1.1 Self-reported clinical reasoning skills – FIT

My systematic review of the literature found that assessing clinical reasoning skills was challenging, as there are few available validated and reliable measures of clinical reasoning skills. Therefore, I took two approaches to measuring clinical reasoning skills, to ensure that I captured clinical reasoning skills as accurately, validly and reliably as possible. I initially selected a standardised self-reported measure of clinical reasoning skills, the Diagnostic Thinking Inventory (DTI) (Bordage, Grant, & Marsden, 1990). The DTI consists of two sub-scales; one measures flexibility of thinking (21 items), the other

measures the structure of knowledge in memory (20 items). Flexibility in thinking refers to the variety of thought processes clinicians use in the diagnostic process; the structure of knowledge in memory, refers to availability of knowledge in memory (Bordage et al., 1990). Flexibility in thinking measures analytical reasoning, in that flexibility in thinking refers to the ability to generate new ideas, understand alternative outcomes and self-reflect. Structure of knowledge in memory reflects non-analytical reasoning processes, which require automatic access to the memory store and content-specific information that has already been stored (Durning et al., 2016).

The DTI has been validated by the authors and it has been shown to detect differences between students and professionals' reasoning (Bordage et al., 1990). The questionnaire has been found to be reliable – the internal consistency of the flexibility sub-scale was acceptable (Cronbach's α = 0.72) and test re-test reliability was also acceptable (Bordage et al., 1990; Round, 1999). To my knowledge it has only been used in two studies to detect change in clinical reasoning skills over time; neither of these studies were eligible for inclusion in my systematic review (Bateman, 2013; Lee et al., 2010). While these studies did not find a significant difference over time or between groups on the DTI, this may have been due to a lack of power to detect an effect, as their sample sizes were ≤ 80. There are potential benefits to using the DTI over other types of assessment, such as assessing reasoning via performance on a patient case. The DTI is a standardised measure that is not dependent on knowledge or affected by variations in the difficulty of patient cases, which may affect scores.

I chose to use only the FIT sub-scale of the DTI because it asks questions on data gathering, consideration of alternative diagnoses and seeking evidence to refute or support hypotheses; all key skills I predicted in my logic model that eCREST might influence. Additionally, feedback from medical students was that the 41 item DTI was too long and could potentially induce survey fatigue. I piloted the FIT sub-scale with experts and students to ensure that the measure had both face and content validity. Both groups agreed that it was representative of the concept but some of the wording was amended to improve clarity. The final version of FIT scale can be found in **Appendix 5.** Higher scores on the FIT sub-scale were indicative of better clinical reasoning skills. Each item was scored on a scale of 1-6; the score can range from 21-126. The FIT was measured at baseline, one week after baseline and one month after baseline. I used the FIT in the feasibility trial to determine the feasibility of using the FIT outcome measure in a definitive full-scale trial.

5.7.3.1.2 *Observed clinical reasoning skills – Key Features*

5.7.3.1.2.1 Developing the key features

As was shown in my systematic review of current online patient simulation evaluations in **Chapter 3**, one of the more commonly used validated outcome measures used by previous trials was the Key Features Problem (Bordage, Brailovsky, Carretier, & Page, 1995; Page et al., 1995). Key Features Problems are based on a clinical problem presented via a patient case and are typically followed by two or three questions. Questions are designed to assess clinical decisions and relate only to the most important steps in resolving the problem. Typical questions would include asking students to list the most likely

conditions and what tests they would order to help them refine their diagnosis. The questions tend to be problem specific so can vary depending on each case. Students are then scored on their answers to these questions. This measure has been found to be both valid and reliable and findings from my systematic review showed the measure was able to detect significant differences in clinical reasoning skills between groups in two previous studies (Bordage et al., 1995; Farmer & Page, 2005).

The Key Features Problems are similar to the patient cases in eCREST; eCREST also presents clinical problems via patient cases and asks students to suggest a differential diagnosis, select questions to ask the patient, request examination results and make a management plan for that patient. Thus, while the eCREST cases were not designed as Key Features Problems, they were based on the same principles and could be used to assess clinical reasoning skills. I used the data collected by eCREST on the patient cases to measure the key features. This complemented data gathered by the FIT survey, as it assessed the observed real-time reasoning of students, which was not captured by the retrospective subjective self-reported measure of reasoning.

I identified the key features for the cases by referring to the logic model for eCREST and the learning objectives of what I expected students to gain from using eCREST. I created key features based on the measurable outcomes that I predicted to occur if the expected mechanisms of change had worked and influenced behaviours and outcomes, as outlined in my logic model in **Chapter 4, Section 4.5**. Therefore, I focused on assessing behaviours and outcomes

regarding data gathering and diagnostic hypothesis testing. I predicted these behaviours and outcomes would be improved by eCREST, as eCREST could address the cognitive biases that can affect reasoning and improve students' mental representations of illnesses, through facilitation of analytical and non-analytical thinking strategies. **Table 5-4** presents my initial ideas for key features and how I intended to assess them using eCREST. Initially I wanted to create a total clinical reasoning score by combining the score from each key feature.

Table 5-4 Initial key features to assess clinical reasoning skills for each patient case

Key Features	Cognitive biases that key feature aims to overcome	How this was measured in the data	Notes
1. Take a focused and relevant history	Unpacking principle (to what extent does the student elicit enough information to make an appropriate decision).	Proportion of all relevant questions and examinations asked out of total questions asked.	
2. Gathers necessary information	Unpacking principle (to what extent does the student elicit enough information to make an appropriate decision).	Proportion of essential questions and examinations asked out of all essential examinations and questions for that patient case.	
3. Adapts a diagnosis in a flexible manner, according to new information	Anchoring (how fixed is the student to their initial diagnosis). Confirmation bias (is the student responding to information that may refute their hypothesis).	a) How many times did student change their diagnosis list or order?	Every student changed their diagnosis at least once.
	Anchoring (how fixed is the student to their initial diagnosis and do they change their mind to more appropriate hypotheses). Confirmation bias (is the student responding to information that may refute their hypothesis).	b) Number of correct diagnoses at initial diagnosis vs final diagnosis.	Subtract final number of correct diagnoses from initial number of correct diagnoses.
4. Prioritise differential diagnoses	Anchoring (how fixed is the student to their initial diagnosis and do they change their mind to more appropriate hypotheses).	Captured in variable above - How many times did they change their diagnosis list or order?	Capturing the activity of prioritising.

5. Consideration of lung cancer/most serious diagnosis in their differential diagnosis	Confirmation bias (has the student only considered the most likely diagnosis and sought information to confirm that hypotheses and ignored rarer possibilities for diagnosis).	a) Binary variable – did the student consider lung cancer at initial differential diagnosis? b) Did the student consider lung cancer at final diagnosis? c) Proportion of students who removed and added lung cancer (or most serious diagnosis) from final diagnosis if had it in their initial?	Relevant to look at this as this study was influenced by the research study that found GPs often did not elicit enough relevant information to appropriately diagnose lung cancer. Additionally, this study has been funded by the Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis, so will be of interest to them.
6. List the most appropriate diagnoses for the patient	N/A but tests knowledge of conditions that can present with respiratory and related symptoms.	d) Proportion of final differential diagnoses which are recommended by GPs.	
7. List the most appropriate tests to order to manage the patient	N/A but tests knowledge of how to manage patients in general practice.	e) Proportion of investigations which are recommended by GPs.	
8. List the most appropriate follow-up for the patient	N/A but tests knowledge of how to follow-up patient in general practice and referral guidelines.	Binary variable – did the student choose the recommended follow-up option?	

5.7.3.1.2.2 Establishing face validity of the key features

In a meeting with my supervisors (JS and APK) and three GP registrars (SM, JH and SG), my suggested key features were discussed and assessed for their face and content validity. In this meeting, it was decided the key features should focus on the thought processes involved in making clinical decisions, including the gathering and interpretation of data, and not on the final consultation outcomes e.g. diagnostic accuracy. This was because the patient cases eCREST presented did not have one correct diagnosis, as they were designed to be cases that exhibited vague and non-specific symptoms that could be indicative of several conditions. Additionally, the GP registrars felt that eCREST was designed to address errors in data gathering and improve flexibility in thinking, by encouraging students to reflect and reconsider their diagnoses but would not necessarily improve diagnostic accuracy. Diagnostic accuracy would depend on good clinical reasoning skills but would also rely on other factors such as knowledge of the relevant conditions, which eCREST does not provide extensive training on. It was agreed that the key features should measure the specific cognitive errors that eCREST aimed to address. Consultation outcomes (key features 3b-8 in **Table 5-4**) were analysed but as separate outcomes to clinical reasoning skills. See consultation outcomes **Section 5.7.3.2.1** for more details.

The key features that clinicians, supervisors and I agreed on are described in **Table 5-5,** along with a description of how they were measured using available data collected from the patient cases in eCREST. It was decided

in the meeting that there was no feasible way of measuring the cognitive error of confirmation bias, as it was not feasible within the timeline of my PhD to track whether students sought information that disproved their hypotheses. This evidence was, however, captured by my Think Aloud study that observed students' reasoning while using eCREST in **Chapter 8**.

Table 5-5 Final key features identified to assess clinical reasoning skills for each patient case

Learning objectives	Key feature number and name	Cognitive bias eCREST aimed to address	How this was measured in the data
Take a focused and relevant history	1. Relevant information gathered.	Unpacking principle – not gathering enough information to make an appropriate diagnosis.	Proportion of all relevant questions and examinations asked out of total questions and examinations asked.
Gathers necessary information	2. Essential information identified.	Unpacking principle – not gathering the most essential information to make an appropriate diagnosis.	Proportion of essential questions and examinations asked out of all essential examinations and questions identified by experts.
Adapts and prioritises a diagnosis in a flexible manner, according to new information	3. Changed diagnosis.	Anchoring – staying fixed upon an initial diagnosis.	How many times did the student change their diagnosis list or order?

To calculate the key features shown above, I recruited two GP registrars in my department (SG and JT) to identify the essential and relevant questions, bedside tests and physical examinations for each patient case. They also identified the appropriate final diagnoses, further tests to manage the patient and follow-up of the patient, so that consultation outcomes could also be measured (see consultation outcomes **Section 5.7.3.2.1**). Cohen's Kappa was calculated to determine the inter-rater agreement, other than expected by chance, between the GP registrars on the key features and consultation outcomes. Agreement on all key features and consultation outcomes was good, all key features had at least 67% agreement and Kappa scores are noted in **Table 5-6** (Viera & Garrett, 2005). The lowest agreement between the GP registrars was for the essential questions. In discussion with the GP registrars it was found this was mostly because one registrar had included some patient perspective questions as essential and one had not. Thus, consensus for a definition of what was deemed an essential question was reached and final decisions made for each patient case. The definition of what constituted an essential item of information (question or test result) was: *a question or test result response that would change the differential diagnosis of the patient case irrespective of the answers to the other questions.* A list of all the essential and relevant questions and physical examinations, the recommended diagnoses, investigations and follow-up for each patient can be found in **Appendix 3 & 6**.

Table 5-6 Inter-rater agreement on the factors needed to create key features from cases 1-4

Factors needed for key feature assessment and consultation outcomes	Average Kappa across case 1-4
Essential questions	0.67
Relevant questions (which includes essential questions)	0.86
Relevant physical examinations and bedside tests	0.88
Recommended diagnoses	0.90
Recommended investigation	0.70
Recommended follow-up	1.00

Unlike Key Features Problems described in the literature, there was not an overall score for the key features. Instead each of the key features was considered a separate outcome and data analysis was undertaken on each feature. This was based on advice from a statistician and GP registrars, as it was uncertain whether each key feature was equally important for clinical reasoning skills. In addition, it would be useful to see which aspects of clinical reasoning skills eCREST may influence. These outcomes were compared between groups on patient case 4, which both the intervention group and control group completed. It was also used to compare clinical reasoning skills within the intervention group, as key feature scores can be compared between patient case 1 and patient case 3, to explore improvement over time. These measurements helped to determine the feasibility of using key features identified in the eCREST patient cases to assess clinical reasoning skills in a definitive full-scale trial. The validity and reliability of the key features were also measured and the data were used to refine the sample size calculation for a definitive full-scale trial in the future.

5.7.3.2 Outcomes related to clinical reasoning skills

5.7.3.2.1 Consultation outcomes

While it was suggested in the meeting with my supervisors and GP registrars to not focus on consultation outcomes, such as whether students identified the appropriate diagnoses, these outcomes were considered of interest to my supervisors and me. Identifying a more accurate diagnosis is an important and positive outcome for students, educators and patients. Additionally, consultation decisions are inextricably related to clinical reasoning skills, as they are the outcome of the clinical reasoning processes of data gathering and interpretation. Indeed, my systematic review revealed that diagnostic accuracy is often used to measure clinical reasoning skills. Therefore, consultation outcomes were measured as a separate but linked measure to clinical reasoning skills.

The consultation outcomes included diagnostic accuracy, and this was measured by calculating the proportion of appropriate diagnoses that students identified. I also captured whether the students identified the most serious possible conditions, such as lung cancer or heart failure (depending on the case), in their differential diagnosis. I measured the proportion of appropriate investigations identified and their selection of the appropriate follow-up to assess their management reasoning. The consultation outcomes were used to determine the predictive validity of the key features, as described above. The consultation outcomes were also compared between the intervention and control group to explore differences between them.

5.7.3.2.2 Knowledge of respiratory medicine - MCQs

Medical knowledge of respiratory symptoms, diagnosis and management was measured, as there is a relationship between knowledge and reasoning (Norman et al., 2016). My logic model discussed in **Chapter 4** describes how knowledge contributes to students' mental representations of illnesses and their thought processes. Additionally, it was useful to have a baseline benchmark of students' specific knowledge of the topics covered in the patient cases presented by eCREST, so that this could be adjusted for in the data analyses if necessary. Knowledge of the medical information related to the patient cases was measured by 20 multiple-choice single best answer questions at registration, and again one month after registering. The questions related to the management and referral of patients presenting with respiratory and related symptoms in general practice. They were developed by a team of GP registrars, in consultation with GPs and medical educators from an advisory group. Students received 12 questions at baseline and a further 12 questions one month after (Time 2). At Time 2 eight of these questions were new questions and four were repeated questions from baseline. The questions were randomised so that each student received the questions in a different order and had different questions repeated to remove any bias that might occur due to varying difficulty of the questions. Each question was worth one point, scores could range from 0-12, with higher scores indicating greater knowledge. The MCQs can be found in **Appendix 7**. Changes in knowledge over time or between groups were not analysed using inferential statistics, as knowledge was not intended to change because of using eCREST but it would be used as

a covariate in the analyses if found to be different between the intervention and control group at baseline.

5.8 Validation of clinical reasoning skills outcomes

I identified in my systematic review that a weakness of previous studies was that they had not attempted to validate the outcome measures they used to assess clinical reasoning skills. Therefore, I validated both the self-reported measure of clinical reasoning skills (FIT survey) and the observed measure of clinical reasoning skills (key features). I have presented the validation results in the methods chapter, so that the results chapter only presents the characteristics of these outcome measures when compared between two groups.

5.8.1 Self-reported clinical reasoning skills (FIT survey)

This measure has been previously validated, as described in **Section 5.7.3**. I sought to establish the construct validity of the measure by comparing it to a relevant measure (knowledge) and the internal consistency of the measure.

5.8.1.1 *Construct validity of the FIT scale*

To assess the construct validity of the FIT I looked to see the correlation between the FIT and outcomes that are known to be related to clinical reasoning that I also measured. Therefore, I looked at the correlation between self-

reported clinical reasoning skills and respiratory knowledge by correlating the FIT survey and the MCQs. **Table 5-7** presents the Spearman's rank coefficient between self-reported clinical reasoning skills and respiratory knowledge at each time point. Self-reported clinical reasoning skills at baseline were significantly positively correlated with respiratory knowledge at baseline ($r_s = 0.13$, $p \leq 0.05$, $n = 240$). Self-reported clinical reasoning skills at Time 1 were also significantly correlated with respiratory knowledge at baseline ($r_s = 0.21$, $p \leq 0.05$, $n = 183$). Self-reported clinical reasoning skills at Time 2 were significantly correlated with respiratory knowledge at baseline ($r_s = 0.21$, $p \leq 0.05$, $n = 140$) and at Time 2 ($r_s = 0.18$, $p \leq 0.05$, $n = 140$). Thus, overall self-reported clinical reasoning skills and respiratory knowledge do appear to be related and the FIT scale can be seen to have good construct validity.

Table 5-7 Correlation matrix showing the relationship between self-reported clinical reasoning skills and respiratory knowledge

	FIT Pre	FIT T1	FIT T2	MCQ Pre	MCQ T2
FIT[1] Pre	1				
FIT T1	0.69*	1			
FIT T2	0.59*	0.66*	1		
MCQ[2] Pre	0.14*	0.20*	0.21*	1	
MCQ T2	0.07	0.02	0.18*	0.25*	1

[1] Flexibility in Think (FIT) scale
[2] Respiratory Knowledge quiz
Note. *$p \leq .05$

5.8.1.2 *Internal consistency of the FIT scale*

The internal consistency of the FIT at baseline was adequate (Cronbach's $\alpha = 0.66$).

5.8.2 Observed clinical reasoning skills

To establish whether the key features were valid I assessed the key feature characteristics of the three patient cases in eCREST that were used as the intervention cases. Students who did the three cases all completed the cases over one week. I hypothesised that a good measure of clinical reasoning skills should:

a) be able to predict whether students were able to identify the most serious diagnosis for each patient case and have predictive validity;

b) be able to detect improvements over time.

First, I present the descriptive statistics of the key features and consultation outcomes to provide context to the validation and show the variations in how students performed across the cases.

5.8.2.1 *Descriptive statistics for the key features for patient cases 1-3*

Table 5-8 reports the descriptive statistics of the key features for the intervention patient cases 1-3. Students were best at taking a relevant history for patient case 2; out of all the questions and examinations students asked, on average 83.0% of them were classed as relevant. Students on average took the

least relevant history for patient case 3, in which the 76.5% of information gathered was deemed relevant. The patient case that students identified the most essential information from was patient case 1; out of all essential information recommended to be gathered for the patient case, students gathered a mean of 69.6% of essential items for case 1. Students were worst at identifying essential information from patient case 2; students identified on average 60.6% of essential information, out of all essential information recommended for that case. Students changed their diagnoses most for patient case 3 than the other cases, ($M = 3.6$ times) and least for patient case 1 ($M = 2.6$ times).

Table 5-8 Descriptive statistics for key features of each case

	Patient case	Trial group	*n*	Mean (*SD*)	Median (*IQR*)
Key Features					
1. Relevant information gathered (%)	1	Intervention	99	77.3 (9.1)	76.5 (12.3)
	2	Intervention	95	83.0 (6.0)	83.8 (7.3)
	3	Intervention	94	76.5 (5.3)	76.5 (6.1)
2. Essential information identified (%)	1	Intervention	99	69.6 (18.2)	70.0 (25)
	2	Intervention	95	69.1 (13.8)	69.2 (23.1)
	3	Intervention	94	63.3 (15.7)	65.5 (20.7)
3. Change diagnosis	1	Intervention	99	2.6 (1.1)	3 (1)
	2	Intervention	95	3.0 (1.0)	3 (2)
	3	Intervention	94	3.6 (1.3)	3.5 (1)

5.8.2.2 Descriptive statistics for the consultation outcomes

Table 5-9 reports the descriptive statistics of the consultation outcomes for each patient case. The number of students who improved on the number of correct diagnoses they had in their final differential diagnosis compared to their initial differential diagnosis was greatest for patient case 3 (n = 76/94, 80.9%) and least for patient case 1 (n = 30/99, 30.3%). The most serious diagnosis students should not have missed was lung cancer for patient case 1, and 3 and heart failure for case 2. Students were best at recognising the most serious diagnosis in their initial differential diagnosis for patient case 1 (lung cancer n = 74/99, 74.8%) and worst for patient case 2 (heart failure n = 12/95, 12.6%). Conversely, students were best at recognising the most serious diagnosis in their final differential diagnosis for patient case 2 (n = 76/95, 80.0%) and worst for patient case 1 (n = 65/99, 65.7%). Students performed best on identifying the recommended diagnoses for patient 3; on average students correctly identified 64.2% of the recommended diagnoses. Students performed worst for identifying recommended diagnoses for patient case 2 (M = 52.6%). Students identified the highest proportion of appropriate investigations, such as chest x-rays, for patient case 2 (M = 80.0% of investigations identified) and lowest proportion for patient case 3 (M = 39.4% of recommended investigations identified). Students understanding of how best to follow-up the patient was greatest for patient case 3 in which 57.5% identified the appropriate follow-up option. The poorest understanding was for patient case 1, in which only 9.1% of students identified the correct follow-up option.

Table 5-9 Descriptive data for the consultation outcomes for each patient

	Patient case	Trial group	*n* (%)	Mean (*SD*)	Median (*IQR*)
Consultation outcomes					
1. Improved number of appropriate diagnoses.	1	Intervention	30 (30.3)		
	2	Intervention	60 (63.2)		
	3	Intervention	76 (80.9)		
2. Selected lung cancer (or heart failure for case 2) in their initial diagnosis[1].	1	Intervention	74 (74.8)		
	2	Intervention	12 (12.6)		
	3	Intervention	40 (42.6)		
3. Selected lung cancer (or heart failure for case 2) in their final diagnosis.	1	Intervention	65 (65.7)		
	2	Intervention	88 (92.6)		
	3	Intervention	70 (74.5)		
4. Percentage of final differential diagnoses that were appropriate.	1	Intervention		62.3 (15.7)	66.7 (16.7)
	2	Intervention		52.6 (11.1)	57 (14.3)
	3	Intervention		64.2 (14.9)	66.7 (16.7)
5. Percentage of investigations that were appropriate.	1[2]	Intervention	78 (78.8)		
	2	Intervention		80 (25.5)	100 (33.3)
	3	Intervention		39.4 (22.4)	33.3 (33.3)
6. Selected appropriate follow-up option.	1	Intervention	9 (9.1)		
	2	Intervention	48 (50.5)		
	3	Intervention	54 (57.5)		

[1] As lung cancer was not considered an appropriate differential for patient case 2 the diagnosis that experts recommended should not be missed was heart failure
[2] Initial key feature was to measure the proportion of appropriate investigations to order but in patient case 1 only one option was appropriate so the variable became binary

5.8.2.3 Predictive validity of key features

I hypothesised that the key features would be related to whether students were able to identify the most serious diagnoses for each patient case. This was analysed by conducting a correlation analysis using Spearman's rank coefficient, as the data for the key features was found to be not normally distributed using the skewness/kurtosis test in STATA. Cases were deleted pairwise not listwise, so that students who were missing a value on one of the variables were not excluded from the entire analysis. **Table 5-10** shows the correlation between identification of the most serious condition for each patient case and the key features. The only significant correlations were for patient case 3, where there was a weak positive correlation between the proportion of essential information identified and selection of lung cancer in their final differential diagnosis ($r_s = 0.31$, $p \leq .01$, $n = 92$). A similar relationship was found between the number of times students changed their diagnosis and selection of lung cancer in their final differential diagnosis for case 3 ($r_s = 0.33$, $p \leq .001$, $n = 94$).

Table 5-10 Correlation between key features and identification of the most serious condition in final diagnosis

	Trial group	Key Feature No.	n	r_s	P value
Patient case 1 (lung cancer diagnosis selected)	Intervention	1	99	-0.001	0.99
		2	99	0.186	0.07
		3	99	0.042	0.68
Patient case 2 (heart failure diagnosis selected)	Intervention	1	93	-0.148	0.16
		2	93	0.027	0.82
		3	95	0.155	0.13
Patient case 3 (lung cancer diagnosis selected)	Intervention	1	92	-0.111	0.29
		2	92	0.308*	0.003
		3	94	0.325*	0.001

5.8.2.4 *Sensitivity of key features to detect improvement*

It was also hypothesised that students may improve their clinical reasoning skills, assessed by the key features, over time. As students could complete the cases in any order I only included participants who completed the cases sequentially from patient case 1 to patient 3. This was to ensure that data were comparable and not affected by the order of the cases. As the data was not normally distributed for the key features, a Wilcoxon signed rank test was conducted to compare the three key features from patient case 1 and patient case 3.

The median proportions for key features one and two for each patient case are shown in **Figure 5-2.** The proportion of relevant information elicited from the patient, out of all the information students gathered, did not significantly

improve from patient 1 (*Mdn* = 0.77, *IQR* = 0.11) to patient 3 (*Mdn* = 0.77, *IQR* = 0.06), *Z* = 0.16, *p* = 0.88, *n* = 76. Moreover, there was no linear relationship, as students appeared to gather more relevant information for patient case 2 than case 1 or case 3. The median proportion of essential information students gathered, out of all possible essential information available, significantly decreased from patient 1 (*Mdn* = 0.70, *IQR* = 0.18) to patient 3 (*Mdn* = 0.66, *IQR* = 0.21), *Z* = 3.83, *p* ≤ .001, *n* = 76. Further exploration of the data revealed that the students asked more essential questions in total over time, the number increased in a linear fashion from case 1 to case 3, but the number of possible essential questions also increased with each case. Therefore, proportionally students asked less essential questions and tests over time.

Figure 5-2 Median proportion of relevant and essential information gathered (key features 1 and 2) for patients 1, 2 and 3
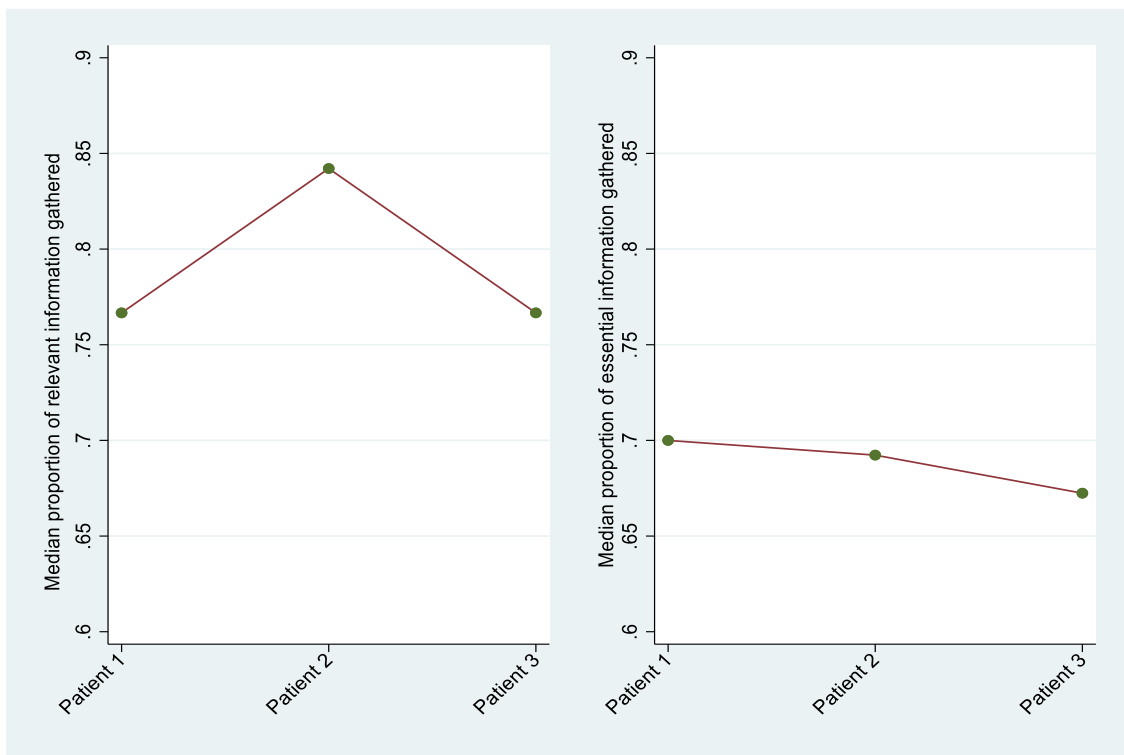
The median number of times students changed their diagnoses significantly increased from patient 1 (*Mdn* = 3.0, *IQR* = 1.0) to patient 3 (*Mdn* = 4.0, *IQR* = 2.0), z = -5.48, $p \leq .001$, n = 78 (**Figure 5-3**).

Figure 5-3 Median number of times students changed diagnoses (key feature 3) scores for patients 1, 2 and 3
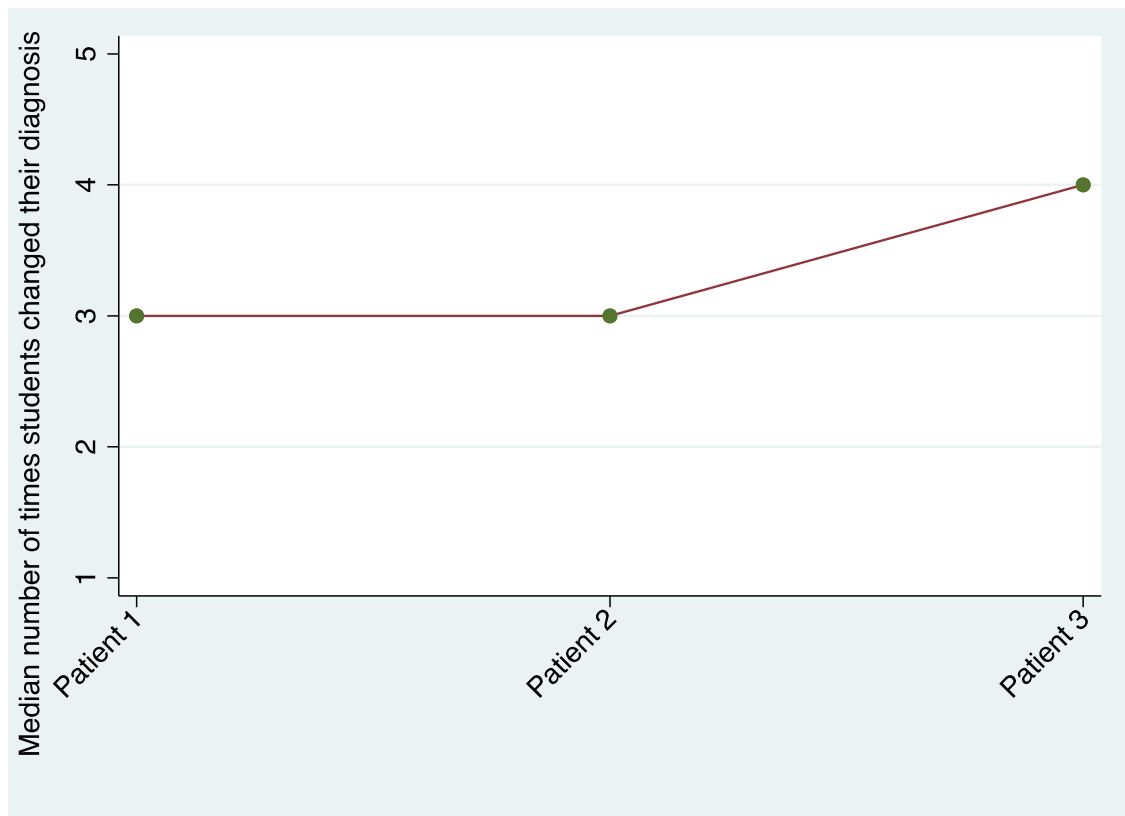
5.8.3 Summary of the validation of the key features

The observed measures of clinical reasoning skills, the key features, had good face and content validity, as they were developed in consultation with experts. However, they may have poor predictive validity, in that good performance on the key features was only partially related to being able to identify serious conditions. It was also unclear how sensitive the key features

were to detecting improvement. There was no improvement from patient case 1 to patient case 3 in the amount of relevant information students gathered and a significant decrease in the amount of essential information they identified. These patterns were not linear and seemed to change depending on the patient case. The cases were designed to be relatively similar in structure and content but there were variations in the cases that made it difficult to compare performance across cases. There were more possible essential questions and tests for students to identify in case three compared to case one (29 versus 20) despite there being a similar number of overall questions (52 versus 51). Students did on average increase the number of questions they asked from case one to case three (24.7 versus 28). However, this increase was not enough to detect all the essential information for case three. Students may have been misled or confused by this difference in the cases, as they may have expected the latter two cases to be the same structure as the first case, in terms of number of relevant and essential questions to identify. They may have felt pressure to stop asking questions based on their experience from case one and felt they were asking too many questions and tests, rather than they did not identify that a question or test was important to ask. The difference in the content of the cases may also have affected the comparability of cases over time. The patients did present with different symptoms from case one to case three and had different recommended differential diagnoses, although all were respiratory and related symptoms and diagnoses that would typically present in primary care. Nevertheless, the variations in symptoms may have affected the level of difficulty of the cases and the students may have had a different level of knowledge of the different symptoms and associated diagnoses presented in

the different patients. Thus, for future validations of the key features it may be important to ensure the structure and content of the cases is similar so that cases are comparable. However, this should not be an issue when comparing the intervention and control group on case 4 because the case was the same for both groups.

Key feature 3, the number of times students changed their diagnosis, did significantly improve over time. This suggested that repeated use of patient cases in eCREST could improve flexibility in thinking about their diagnosis. Furthermore, this key feature was less likely to be affected by variations in the number of essential information to identify in the case, so it may be more appropriate to assess across cases. However, it could still be affected by level of knowledge and difficulty in a case, as students may change their mind more in difficult cases or if they know less about a diagnosis.

5.9 Statistical analysis

All analyses were undertaken using statistical analyses package STATA version 15 (StataCorp, 2017). The level that was considered statistically significant was 5%. The analyses are presented in four sections in **Chapter 6**: section one presents the analysis of the primary outcomes of feasibility; section presents two presents the acceptability of eCREST; section three presents the demographics and baseline comparisons and section four presents the potential impacts of eCREST on clinical reasoning skills, using the FIT survey and the key features.

5.9.1 Analysis of feasibility and acceptability outcomes

5.9.1.1 Feasibility

5.9.1.1.1 Uptake and follow-up

I presented descriptive data showing the number of eligible students, the number and percentage of participants who registered, out of all eligible and the number and percentage of participants who completed the follow-up at Time 1 and 2, out of those registered. Data was presented via a participant flow diagram accounting for all participants. I compared percentages of students who were recruited and follow-up rates between trial group, universities and cohorts using chi-square tests.

5.9.1.1.2 Recruitment

Recruitment was analysed by reporting the descriptive statistics of the number and proportion of different ways in which students heard about this study, and their motivation for taking part in the study, out of all those who registered. I also described the different methods of recruitment for each cohort in more detail and provided the number of students who were recruited at each site, for each time point. A binary variable was created to reflect whether students were in cohort 1 or 2. I then summarised the factors that I found were related to recruitment success in this feasibility RCT.

5.9.1.1.3 Completion

I presented the number and percentage of students who completed the three patient cases out of those in the intervention. Completion was defined as the submission of a final differential diagnosis for the patient.

5.9.1.1.4 Duration of the intervention

I presented the median (and inter-quartile range) of the time taken to complete each case, as it was assumed that the data would not be normally distributed given students had up to a week to complete the three cases and timings would vary widely.

5.9.1.2 Acceptability

Acceptability of eCREST was analysed by reporting the number and proportion of students who agreed or strongly agreed, neither agreed or disagreed, or disagreed or strongly disagreed, for questions 1-6 on the acceptability survey. I also compared the acceptability between the universities and cohorts using a chi square for each of the six questions on the survey. Students' free text comments were analysed by counting and comparing the number of occurrences of different comments and suggestions using summative content analysis (Hsieh & Shannon, 2005).

5.9.2 Participant characteristics

Data on the total number of participants by gender, age, university, cohort and previous degree was given to provide an overview of the sample. To verify that the randomisation of participants did distribute characteristics evenly between groups and reduce bias, I compared baseline characteristics of the intervention and control group by comparing mean scores on the MCQs and the FIT at baseline, using one-way ANOVA or Man-Whitney test, depending on normality of the distribution of the data. I compared gender, age, university and

cohorts between intervention and control groups, using a chi square. I also compared the same demographic and baseline variables between universities and between cohorts to check for any differences in these characteristics between the samples, using one-way ANOVAs or Mann-Whitney tests depending on normality.

5.9.3   Analysis of clinical reasoning skills outcomes

5.9.3.1 Self-reported clinical reasoning skills - FIT

I present the data on the normality of the self-reported clinical reasoning skills survey, which is visually represented via Quantile-Quantile plots (Q-Q plots), histograms and boxplots (to highlight outliers). Sensitivity analysis was conducted, where outliers were present, in which the analysis would be repeated with outliers removed. I used a mixed factorial ANOVA to assess whether clinical reasoning skills improve over time and between groups, and assess interactions between time and group allocation.

5.9.3.2 Observed clinical reasoning skills and consultation outcomes

Logistic regression analyses were used to determine if there was a relationship between group allocation and clinical reasoning skills, as measured by the key features. A separate logistic regression analysis was conducted for each key feature, which was the outcome variable in the model, and group allocation was the predictor variable in each model. Baseline self-reported clinical reasoning skills and respiratory knowledge were not controlled for because they introduced multicollinearity into the models. Furthermore, as there

were no significant differences in these variables between the intervention and control group at baseline (see **Chapter 6**), it was not necessary to control for these factors as they were unlikely to influence the models. Indeed, sensitivity analysis revealed that the exclusion of baseline self-reported clinical reasoning skills and knowledge at baseline did not change the significance of the associations in any of the models. The same analyses were conducted for consultation outcomes.

For the key features and consultation outcomes data that were proportions, such as the proportion of essential information gathered, linear regressions were inappropriate to use because the values were bounded by 0 and 1 and the assumption of equal variance was not satisfied. For the key features and consultation outcomes where data was a proportion, a logistic regression was performed. The effect of calculating the log odds of the outcome was to extend the ends of the distribution to create a nearly linear relationship (Dixon, 2008; Warton & Hui, 2011). This was carried out in STATA using the generalised linear models (glm) function, in which a logit model and a binomial distribution was specified. The key features or consultation outcomes that were not a proportion, were binary variables or categorical variables, and a logistic regression or multinomial logistic analysis was carried out for these outcomes. Studentised residuals were checked to ensure there were no outliers that may have influenced the data; if residuals of >2 or <-2 were found the data were checked.

5.9.4 Treatment of missing data

Analysis was only completed when data was available for all the variables used in each analysis; no multiple imputation was used. Therefore, those students who had missing data at any time point were excluded from analysis.

5.10 Summary

This chapter described how I evaluated an online patient simulation tool, eCREST, by undertaking a feasibility RCT to estimate feasibility, acceptability and the potential impacts of an online patient simulation on clinical reasoning skills. I described how I used two different measures to assess clinical reasoning: a self-reported survey to detect general use of thinking strategies and observational data that was automatically collected by eCREST to provide insight into students' behaviours and consultation outcomes, such as data gathering and diagnostic hypothesis testing. The validity and reliability of the measures was tested. While the key features were found to have limited validity and reliability, they were used to assess reasoning because they provided valuable insight into students' reasoning, which could not otherwise have been obtained. **Chapter 6** will describe the results of the feasibility RCT.

6 Results: Feasibility RCT

In this chapter, I present the results of the feasibility RCT. Section one, presents the outcome relating to the feasibility of conducting an RCT. Section two, presents the acceptability of eCREST. Section three, presents participants' characteristics and comparisons of baseline and demographic characteristics by trial group, university and cohort. Section four, presents the exploration of the potential impact of eCREST on clinical reasoning skills.

6.1 Section one: The feasibility of conducting an RCT

6.1.1 Uptake and follow-up

A participant flow diagram in presented in **Figure 6-1** in line with the consolidated standards of reporting trials statement (CONSORT; Schulz, Altman, & Moher, 2010). The figure shows the number of students who were eligible and the number of students who completed the follow-up evaluations at both time points by trial group. Of the 1,454 eligible medical students across the three UK medical schools included in this study, 264 participated (18.2%). Of those that participated, 70% (n = 185/264) stayed in the study one week after baseline and 53% (n = 140/264) stayed in the study one month after baseline. Results from chi-square tests showed that there was not a significant difference, between the intervention and control groups, in the proportion of students who stayed in the study one week after baseline (72% and 68% respectively) (χ^2 (1) = 0.65, p = 0.42) or after one month (57% and 55% respectively) (χ^2 (1) = 0.34, p = 0.56).

Figure 6-1 CONSORT flow diagram of the progress through the phases of the feasibility RCT of the intervention and control group
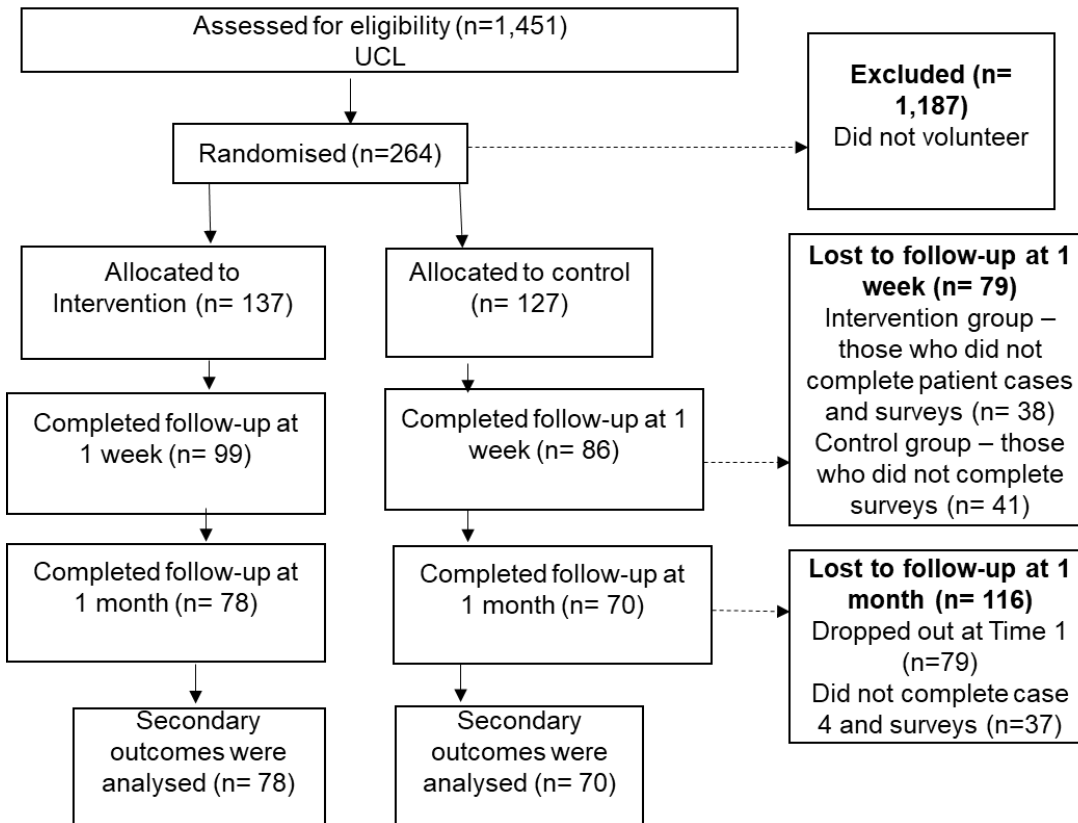
Table 6-1 displays data on uptake and retention at the different medical schools that I recruited from. Uptake was similar in SB ($n = 136/264$, 52%) and SA ($n = 112/264$, 42%) and considerably lower at SC ($n = 16/264$, 6%). There was no significant difference in the proportion of students at each university who stayed in the study one week after baseline. However, there was a significant difference in the proportion of students at each university who stayed in the study one month after baseline, $\chi^2 (2) = 9.58$, $p \leq 0.01$. Students at SA appeared to be more likely to drop out of the study after one month ($n = 47/112$, 42%)

than those at SB (n = 83/136, 61%) and SC (n = 10/16, 63%). Uptake also

differed considerably between cohorts with just under 70% (n = 183/264) taking

part in cohort 2 and 31% (n = 81/264) taking part in cohort 1. Those in cohort 1

were significantly less likely to stay in the study one-week post baseline (n =

45/81, 56%) than those in cohort 2 (n = 140/183, 77%), χ^2 (1) = 11.75, $p \leq 0.001$.

This was also true one month post baseline (n = 29/81, 36% and n = 111/183,

61% respectively), χ^2 (1) = 13.92, $p \leq 0.001$.

Table 6-1 Comparison of uptake and retention in the study by university and cohort

		No.	P Value[1]
Registered to take part			
	University		
	SB	136	
	SA	112	
	SC	16	
	Cohort		
	1	81	
	2	183	
Completed follow-up at Time 1			
	University		
	SB	98/136 (72%)	
	SA	74/112 (66%)	0.36
	SC	13/16 (81%)	
	Cohort		
	1	45/81 (56%)	≤ 0.001[2]
	2	140/183 (77%)	
Completed follow-up at Time 2			
	University		
	SB	83/136 (61%)	
	SA	47/112 (42%)	≤ 0.01[3]
	SC	10/16 (63%)	
	Cohort		
	1	29/81 (36%)	≤ 0.001[4]
	2	111/183 (61%)	

[1] P value taken from chi-square test
[2] χ^2 = (1) = 11.75
[3] χ^2 = (2) = 9.58
[4] χ^2 = (1) = 13.92

6.1.2 Recruitment

Most students were recruited to the study via an email from their respective medical school requesting participation (n = 173/264, 66%), as shown in **Figure 6-2**. Most students reported they took part in the study to help them prepare for clinical practice (n = 39/98, 40%). A similar proportion of students reported that they took part to be a better doctor (n = 32/98, 33%) and to receive a voucher (n = 27/98, 27%).

Figure 6-2 Students' reports of how they were recruited
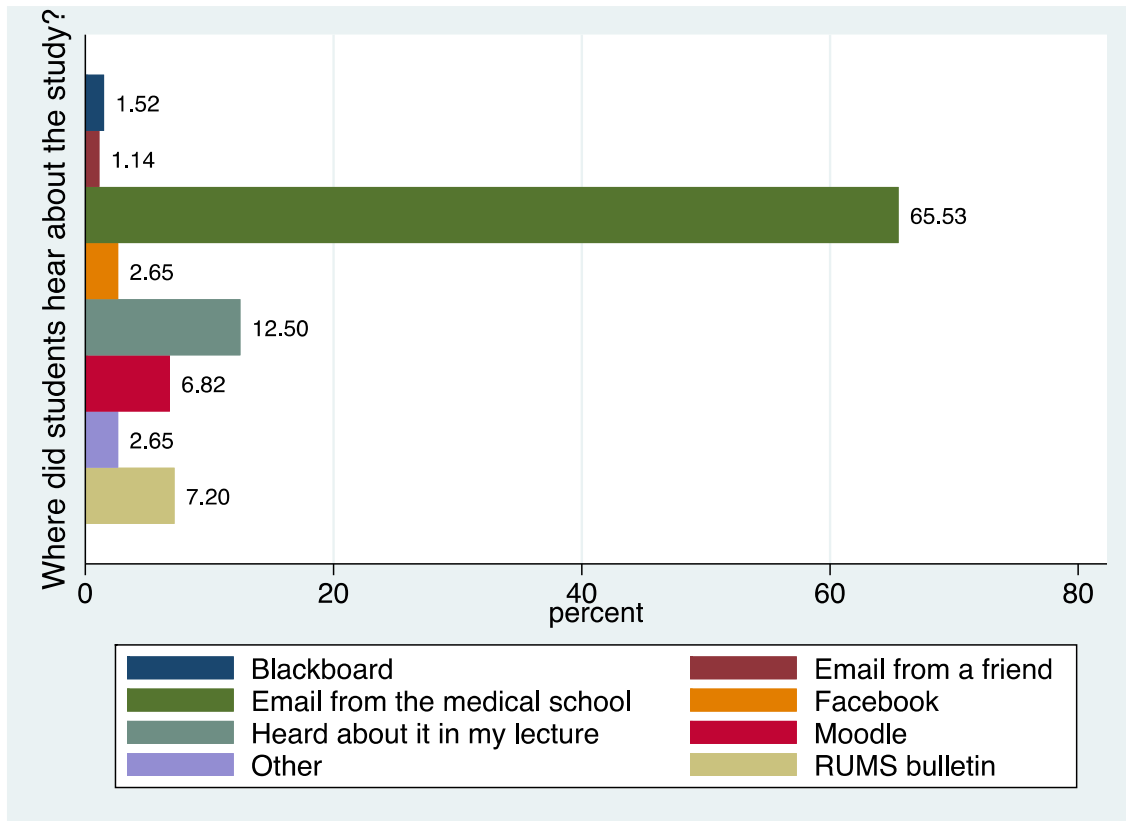
6.1.2.1 Recruitment of cohort 1

In discussion with the medical school tutors, it was initially thought that uptake to the study would be greater if eCREST was advertised to students on

their teaching day and asked to take part in their lunch hour in a computer room on their campus. However, this method proved ineffective in the spring term of 2017, as only 10 students were recruited at one site. Recruitment at both sites improved in the summer term of 2017, most likely because students had completed their examinations and electives, and because I improved the advertisement of eCREST on their teaching day and via social media. Details of how students were recruited, and number of students recruited in cohort 1 is shown in **Table 6-2**.

Table 6-2 Participants recruited for cohort 1

Medical school	Year group	Eligible number	Final exams completed	Month recruited	How were they recruited?	eCREST delivery	Number registered
Academic year 2016/2017							
SA	6	336	Yes	April 2017	Advertised on teaching day and prior to teaching day via news bulletin.	Came to computer room in lunch hour on teaching day.	0
SA	6	336	Yes	June 2017	Advertised on teaching day (with GP registrar) and prior to teaching day via news bulletin.	Came to computer room in lunch hour on teaching day and took part at home part remotely.	44
SB	5	285	Yes	March 2017	Advertised prior to teaching day via email.	Came to computer room after teaching.	10[1]
SB	5	285	Yes	June 2017	Advertised prior to teaching day via email.	Took part at home part remotely.	39

[1] This group were later removed from main analysis because they followed a slightly different procedure.
The procedure was changed for June 2017 groups. See Methods section 5.4.4.1 on procedure for more details.

6.1.2.2 Recruitment for cohort 2

As uptake had been lower than expected in cohort 1, I re-advertised and re-ran the study with SB, SC and SA in the autumn/winter terms of 2017/2018. SB and SC students were recruited via advertisements of eCREST on their eLearning platforms and eCREST was verbally promoted by their tutors in lectures when possible. These students had the opportunity to take part at any time during October 2017-February 2018 and, consequently, they could be undertaking different modules while they were completing eCREST. SA students were recruited in the same way, by advertising the study via their eLearning platform, but students at this site could only take part when they were on their GP placement. It was hoped this would encourage participation in the study, as it would be integrated into a relevant module in their curriculum and would be directly related to their current clinical placement. A brief information pamphlet was provided for their GP tutors to help them discuss the cases and reinforce the learning. Further details of how students were recruited from cohort 2 and numbers recruited are shown in **Table 6-3**.

Table 6-3 Participants recruited for cohort 2

Medical school	Year group	Eligible number	Final exams completed	Month recruited	How were they recruited?	eCREST delivery	Number registered
Academic year 2017/2018							
SA	6	360	No	October-February 2017/2018	Advertised prior, during and at the end of GP placement via Moodle.	Took part at home part remotely.	60
SB	5	325	No	October-February 2017/2018	Students emailed by medical school and advert was placed on Moodle. President of society emailed advert and posted on social media and sent to year 5 representatives.	Took part at home part remotely.	102
SC	5	148	No	October-February 2017/2018	Students emailed by medical school and advert placed on Blackboard. Verbally promoted on teaching days. Advert placed on twitter by course lead and student representative group.	Took part at home part remotely.	16

Overall recruitment in cohort 2 was more successful than cohort 1 at SA and SB. A summary of the lessons learnt on how to recruit students to a similar type of trial can be seen in **Table 6-4** and **Table 6-5.**

Table 6-4 Factors relating to recruitment success

Factors relating to recruitment success	Recommendations from the eCREST trial
Time of year	Providing eCREST to students before final examinations but also around time of revision. Providing after elective when preparing for practice was also an effective strategy but less so than before exams.
Promoted by tutors	Ensuring a key member of staff e.g. head of year or tutor, who students have frequent contact with, promoted the study at any opportunity they could.
Promoted online	Ensuring the study was promoted on the medical school eLearning platforms where students access all their other online learning e.g. Moodle.
Integrated into curriculum	Advertising the study in a module that was relevant e.g. a GP placement module or respiratory module.
Promoted by student representatives	Ensuring the study was promoted by student representatives either via email or social media.
Eye-catching advertisement	Producing eye-catching infographics and presentations describing eCREST, the trial and why it was useful for students to take part.
Promoted by clinician	Acquiring a clinical champion (e.g. GP registrar) who can help to endorse eCREST verbally in lectures or via email or eLearning platforms.
Remote access	Allowing students to take part remotely in their own time. This seemed to be more effective than approaching students on teaching days.
Incentives	Giving students a monetary incentive to complete both follow-ups, which was important given they were volunteers.

Table 6-5 Factors relating to recruitment challenges

Factors relating to recruitment failure	Recommendations from the eCREST trial
Promotion via online newsletters	At SA, the school's newsletter was initially used to promote the study, as the faculty did not send email invitations for research. I found direct emails to students and promotion on their eLearning platforms were more successful, possibly because students were more likely to read those invitations.
Lack of student champions	My approach to recruitment was perhaps too top down, as I sought to invite students through their eLearning platforms or through emails via the medical school. More students may have taken part if I had consulted with students early in the project to discover the best way to advertise research to them and created relationships with student champions who could have helped to promote the study.
Technical failures	On one occasion when students were recruited to take part in the study on their teaching day, the website registration page was not working properly making it difficult for some users to register, which likely resulted in lower uptake.
Lack of information	Initially when students were recruited there was no website or point of information students could refer to to get more information, other than a brief advert. During the study a Facebook page was created to provide more information.
Disconnect with the medical school	Initially the project was advertised as external to the medical school. Other approaches were taken to recruitment later in the project, in which the study was advertised internally by the faculty via email, face-to-face or their learning platform. I feel this may have added more legitimacy to the study and encouraged students to see eCREST as valuable learning.

6.1.3 Completion

Of those that were allocated to the intervention group, who received the three cases, almost three-quarters ($n = 99/137$) completed patient case one in

eCREST. This reduced to just under 70% for the second and third patient cases

(n = 95/137 and n = 94/137 respectively).

6.1.4 Duration of the intervention

The skewness/kurtosis test showed that the duration of each case was not normally distributed, so the median and interquartile ranges (*IQR*s) of each measure were presented. **Table 6-6** shows the median duration in minutes it took for students to complete each case. Students in the intervention group, who completed cases 1-3, had one week to complete all three cases. The median time students spent on cases decreased from case 1 (*Mdn* = 13.3 mins, *IQR* = 11.9) to case 3 (*Mdn* = 12.0 mins, *IQR* = 7.6) but had a wide distribution.

Table 6-6 Time students spent on each case in the intervention

Patient case	No.	Median (*IQR*)	Min	Max
Case 1	99	13.3 mins (11.9)	1.6 mins	2.9 days
Case 2	95	12.7 mins (7.9)	1 min	2.9 days
Case 3	94	12.0 mins (7.6)	0.45 mins	19.2 hours

6.2 Section two: Acceptability

6.2.1 Perceptions of usability and learning

Table 6-7 presents self-reported data from medical students on both the usability of eCREST and their perceived learning from eCREST; 98 students responded to this survey. Almost all students agreed that eCREST was easy to navigate through ($n = 96/98$, 98%) and that the level of difficultly was appropriate ($n = 95/98$, 97%). Most students agreed that eCREST should be used to supplement traditional teaching ($n = 88/98$, 90%). Over 85% of students ($n = 84/98$, 86%) reported that overall eCREST had enhanced their learning and that they would use it again without an incentive. Most students also agreed that eCREST helped them to learn clinical reasoning skills to apply to their clinical work ($n = 80/98$, 82%).

Table 6-7 Medical students reports of the usability of eCREST and self-reported learning from eCREST

Statement	Strongly agree/ Agree n (%)	Neither agree or disagree n (%)	Strongly disagree/ Disagree n (%)	Total N
It was easy to navigate through eCREST	96/98 (98)	1/98 (1)	1/98 (1)	98
The level of difficulty of the material was appropriate	95/98 (97)	3/98 (3)	0/98 (0)	98
eCREST should be used to supplement traditional teaching	88/98 (90)	9/98 (9)	1/98 (1)	98
eCREST helped me to learn clinical reasoning skills to apply to clinical work	80/98 (82)	15/98 (15)	3/98 (3)	98
Overall, using eCREST enhanced my learning	84/98 (86)	13/98 (13)	1/98 (1)	98
I would use eCREST in the future without an incentive	84/98 (86)	10/98 (10)	4/98 (4)	98

Chi square tests revealed there were no significant differences between universities on any of the six statements. However, there was a significant difference between cohorts on students' reports of whether eCREST helped them to learn clinical reasoning skills. Those in cohort 2 were significantly more likely to report that they strongly agreed or agreed (87.7%) that eCREST helped to improve their clinical reasoning skills than cohort 1 (64.0%), χ^2 (2) = 7.5, n = 98, $p \leq .05$). Furthermore, those in cohort 2 were significantly more likely to report that they strongly agreed or agreed (93.2%) that eCREST enhanced their overall learning than cohort 1 (64.0%), χ^2 (2) = 13.7, n = 98, $p \leq .001$). Cohort 2 were also significantly more likely to report that they strongly agreed or agreed (97.3%) that they would use eCREST again without an incentive compared to cohort 1 (52.0%), χ^2 (2) = 31.8, n = 98, $p \leq. 001$). There were no significant differences between cohorts on any of the other statements.

6.2.2 Comments and suggestions for improvement

In total 43 students (n = 43/98, 44%) made comments or suggestions for the improvement of eCREST (**Table 6-8**), for further comments see **Appendix 8**. Several students left positive comments about eCREST, suggesting they found it useful and appropriate for their level of study (n = 9/43, 21%). Moreover, some students also felt that eCREST would be a useful addition to current teaching at the medical school and useful for revision. Importantly, some suggested that eCREST had a positive impact on their clinical reasoning skills and clinical practice, as it helped them to piece together information and think like a clinician.

Four students (9.3%) also commented that while they were motivated to take part because of a voucher they would participate again without an incentive if it were part of their curriculum.

The most common suggestions for eCREST from students were around making eCREST more user friendly ($n = 14/43$, 32.6%). Students suggested they would like more help keeping track of the questions they asked, that the videos could be shorter in length and that the instructions on how to use eCREST could be clearer. The next most common suggestions were to add more patient cases ($n = 11/43$, 25.6%). In particular, they wanted cases in other clinical settings, such as emergency care, and that had a range of difficulty. They also suggested that allowing for more user input in the patient cases would improve eCREST ($n = 5/43$, 11.6%). For example, they suggested changing the format of the questions, investigations and diagnoses from a predefined list to an open-text format. Some students would have preferred to write their own questions, investigations or diagnoses and then view the predefined list. They felt this would increase the difficulty of the task, as it would not provide clues to the appropriate questions, investigations or diagnoses. Students also reported that they would have liked more feedback ($n = 4/43$, 9.3%), particularly on the essential questions they should have asked and the management plans for the patient.

Table 6-8 Medical students' comments and suggestions of improvements for eCREST

Comment	n	Sample of representative quotes
eCREST was appropriate and applicable to clinical practice	9	"Very well designed program. One of the few that actually felt helpful in improving my clinical practice." (Student A) "It allowed me to train my thoughts to think like a clinician and showed the way one should put together signs and symptoms." (Student B)
Motivated to use eCREST again without a voucher	4	"I did it for a voucher....but it was actually so informative. Would do it without a voucher if it was part of the curriculum." (Student C)
Wanted more user input	5	"Before giving a dropdown list of possible differentials, have an open-ended box for the user to write down his/her own differential list. The user can then compare his/her differential list to those in the dropdown list at the end of the exercise." (Student D)
Wanted more feedback	4	"The discussion of the various possible diagnoses was good but would have been useful to have more feedback on choice of investigations and management as this is what I struggle with at this stage more than the diagnosis itself." (Student E)
Wanted more/different patient cases	11	"Design a range of difficulties, some with ambiguous diagnosis and some straightforward." (Student F) "Develop scenarios to test acute/emergency care management." (Student G)
More user friendly	14	"Better way of keeping track of how many questions you have asked, so you don't forget and then are suddenly told to move to next step. This would mean that students pay more attention to the questions they are asking and have left to ask." (Student H)

6.3 Section three: Participant characteristics

The overall demographic characteristics of participants in the study are shown in **Table 6-9**. Of the 264 students who registered to take part in the eCREST trial, there were slightly more male participants (*n* = 142, 54%) than female. This was shown to be similar to the proportion of males and females in their final year at SA (50% male) and SB (51%) but this study had considerably more males than the proportion of males at SC (35%).[1]

In this study, most participants were aged between 23 to 24 years of age (*n* = 152, 58%), which fits with the expected age at which students in the UK would be in their final year of medical school. In this study, 86% were aged 20-26, which was a similar proportion to SC students[2], where 82% of students were aged 21-25. However, SA's demographic was slightly younger with 95% of students aged between 20-26 years and SB's demographic was older with only 38% of students aged 20-24. Most of the students took part in the trial in the second phase of recruitment, or cohort 2 (*n* = 183, 69%) which occurred in the autumn and winter terms of the 2017/2018 academic year. Of the three sites I recruited from, most students were from SB (*n* = 136, 52%). Most students had a previous degree or an intercalated degree, in which they completed a year of study in another field (*n* = 212, 80%).

---

[1] Data provided by SA, SB and SC medical schools.
[2] I was not able to obtain estimates for the exact age range of 20-26 but used the most similar age bracket I was provided data with by the medical schools.

Table 6-9 Overall demographic characteristics of medical students in eCREST trial

Demographic variables	*n*	%
Gender		
Female	122	46
Male	142	54
Age group (years)		
20-22	5	2
23-24	152	57
25-26	68	26
27-28	21	8
>29	18	7
Cohort		
1	81	31
2	183	69
University		
SB	136	52
SA	112	42
SC	16	6
Previous Degrees		
None	15	6
iBSc/ IBMedSci/ BSc/ BA	212	80
MSc/ MA/ MRes	12	5
BDS/ BEng	3	1
Missing	22	8
Total	264	

6.3.1 Comparison of baseline and demographic characteristics by trial group, university and cohort

6.3.1.1 *Differences in demographic and baseline characteristics between intervention and control group*

Table 6-10 presents data demonstrating that the randomisation of participants successfully randomly distributed demographic characteristics across the intervention and control group, as there were no significant differences in age, gender, university or cohort between groups. **Table 6-11** demonstrates that there were no significant

differences between the intervention and control group in baseline self-reported clinical reasoning skills and respiratory knowledge.

6.3.1.2 Differences in demographic and baseline characteristics between universities and cohorts

There were significant differences in the age of students between the universities (χ^2 (2) = 26.7, $p \leq$. 001) (**Table 6-10**). Students at SB tended to be older than those at SA and SC, with 47.8% of SB students being aged 25 or over, compared to 34.1% of SA students and 25.1% of SC students. There were no significant differences between universities in terms of gender.

Significance testing was not conducted to compare cohort allocation by university because SC only participated in cohort 2, which may have biased the results. There was a significant association between age and cohort (χ^2 (4) = 20.1, $p \leq$.001) (**Table 6-10**), with those in cohort 1 being older than those in cohort 2 but this is unsurprising given they took part later in the academic year. There was no significant association between gender and cohort.

The data were found to be relatively normally distributed (see **Section 6.4.1**). A one-way ANOVA revealed that there was no significant difference in baseline self-reported clinical reasoning skills between universities (F (2, 237) = 1.53, p = 0.22) (**Table 6-11**). However, there was a significant difference between the universities in their baseline self-

reported knowledge of respiratory medicine (χ^2 (2) = 26.1, $p \leq .001$), with SA ($M = 9.9$, $SD = 1.3$) scoring higher on average than SC ($M = 9.4$, $SD = 1.7$) and SB ($M = 8.7$, $SD = 1.8$). There was not a significant relationship between cohort and baseline self-reported clinical reasoning skills or knowledge of respiratory medicine (**Table 6-11**).

Table 6-10 Demographic characteristics of medical students in eCREST trial by group allocation, university and cohort

Demographics	Intervention group	Control group		SB	SA	SC		Cohort 1	Cohort 2	
Demographic variables	*n* (%)	*n* (%)	*P* value[1]	*n* (%)	*n* (%)	*n* (%)	*P* value[2]	*n* (%)	*n* (%)	*P* value[3]
Age (Years)										
20-22	4 (2.9)	1 (0.8)		2 (1.5)	1 (0.9)	2 (12.5)		1 (1.2)	4 (2.2)	
23-24	73 (53.3)	79 (62.2)		69 (50.7)	73 (65.2)	10 (62.5)		31 (38.3)	121 (66.1)	
25-26	39 (28.5)	29 (22.8)		34 (25.0)	31 (27.7)	3 (18.8)		33 (40.7)	35 (19.1)	
27-28	11 (8.0)	10 (7.9)		18 (13.2)	2 (1.8)	1 (6.3)		8 (9.9)	13 (7.1)	
>29	10 (7.3)	8 (6.3)	0.49	13 (9.6)	5 (4.5)	0 (0.0)	≤.001*	8 (9.9)	10 (5.5)	≤.001*
Gender										
Female	64 (46.7)	58 (45.7)		57 (41.9)	54 (48.2)	11 (68.8)		31 (38.3)	91 (49.7)	
Male	73 (53.3)	69 (54.3)	0.87	79 (58.1)	58 (51.8)[1]	5 (31.3)	0.11	50 (61.7)	92 (50.3)	0.09
University										
SB	70 (51.1)	66 (52.0)		n/a	n/a	n/a		n/a	n/a	
SA	58 (42.3)	54 (42.5)		n/a	n/a	n/a		n/a	n/a	
SC	9 (6.6)	7 (5.5)		n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cohort										
1	41 (29.9)	40 (31.5)		37 (27.2)	44 (39.3)	0 (0)		n/a	n/a	
2	96 (70.1)	87 (68.5)	0.78	99 (72.8)	68 (60.7)	16 (100)	n/a[4]	n/a	n/a	n/a
Total	137 (51.9)	127 (48.1)		136 (51.5)	112 (42.4)	16 (6.1)		81 (30.7)	183 (69.3)	

[1] *P* value taken from chi-square test
[2] *P* value taken from chi-square test
[3] *P* value taken from chi-square test
[4] Significance testing not done to compare universities and cohorts because some universities only took part in one cohort which would have biased significance test
* *P* value significant

Table 6-11 Baseline self-reported clinical reasoning skills and knowledge by group allocation, university and cohort

Continuous variables			*n*	Mean (*SD*)	Median (*IQR*)	*P* Value
Self-report clinical reasoning skills	Trial group	Intervention	122	83.1 (9.6)	83 (12)	
		Control	118	83.5 (8.8)	84 (11)	
						0.75[1]
	University	SB	125	84 (10.3)	84 (13)	
		SA	100	82.1 (7.9)	83 (10)	
		SC	15	85.3 (7.4)	84 (9)	
						0.22[2]
	Cohort	1	74	84.6 (7.8)	84.5 (12)	
		2	166	82.7 (9.8)	83 (11)	
						0.14[3]
Self-report knowledge of respiratory medicine	Trial group	Intervention	125	9.2 (1.8)	9 (2)	
		Control	126	9.3 (1.6)	9 (3)	
						0.88[4]
	University	SB	130	8.7 (1.8)	9 (2)	
		SA	105	9.9 (1.3)	10 (2)	
		SC	16	9.4 (1.7)	9.5 (3)	
						≤.01*
	Cohort	1	77	9.6 (1.4)	10 (1)	
		2	174	9.1 (1.8)	9 (3)	
						0.07[5]

[1] *P* value taken from t-test
[2] *P* value taken from one-way ANOVA
[3] *P* value taken from one-Way ANOVA
[4] *P* value taken from Mann-Whitney test
[5] *P* value taken from Kruskal-Wallis test
* *P* value significant

6.4 Section four: The exploration of the potential impacts of eCREST on clinical reasoning skills

6.4.1 Self-reported clinical reasoning skills

6.4.1.1 FIT scale descriptive statistics

The intervention group had non-significantly higher self-reported clinical reasoning skills (M = 84.1, SD = 10.3) than the control group (M = 82.4, SD = 9.0) at Time 1, t (183) = -1.14, p = 0.26. The intervention group also had non-significantly higher clinical reasoning skills (M = 84.4, SD = 9.8) than the control group (M = 82.0, SD = 9.4) at Time 2, t (138) = -1.46, p = 0.15 (**Table 6-12**).

Table 6-12 Differences in self-reported clinical reasoning scores at each time point by trial group

FIT Score	Trial group	n	Mean (SD)	Median (IQR)	P value[1]
Baseline	Intervention	122	83.1 (9.6)	83 (12)	0.75
	Control	118	83.5 (8.8)	84 (11)	
Time 1	Intervention	99	84.1 (10.3)	84 (15)	0.26
	Control	86	82.4 (9.0)	83 (13)	
Time 2	Intervention	75	84.4 (9.8)	84 (14)	0.15
	Control	65	82.0 (9.4)	83 (14)	

[1] P values taken from t-test

6.4.1.2 Testing the assumptions for a mixed-factorial ANOVA

6.4.1.2.1 Normality of the distribution

I used Q-Q plots, boxplots and histograms to display the normality of the distributions and to highlight outliers and skewness of self-reported clinical reasoning skills at each time point for the intervention and control group (**Figure**

6-3- Figure 6-5). The self-reported clinical reasoning skills scores at each time point were approximately normally distributed for both groups. The self-reported clinical reasoning skills at baseline and Time 1 were slightly skewed to the right, indicating students were more likely to report higher clinical reasoning skills at baseline and Time 1. The Q-Q plots and boxplots indicated there were four outliers for self-reported clinical reasoning skills at baseline and one of these cases was also an outlier at Time 1; no outliers were identified at Time 2.

Figure 6-3 Q-Q plot, histogram and boxplot showing the normality of the distribution of FIT scores at baseline by group allocation
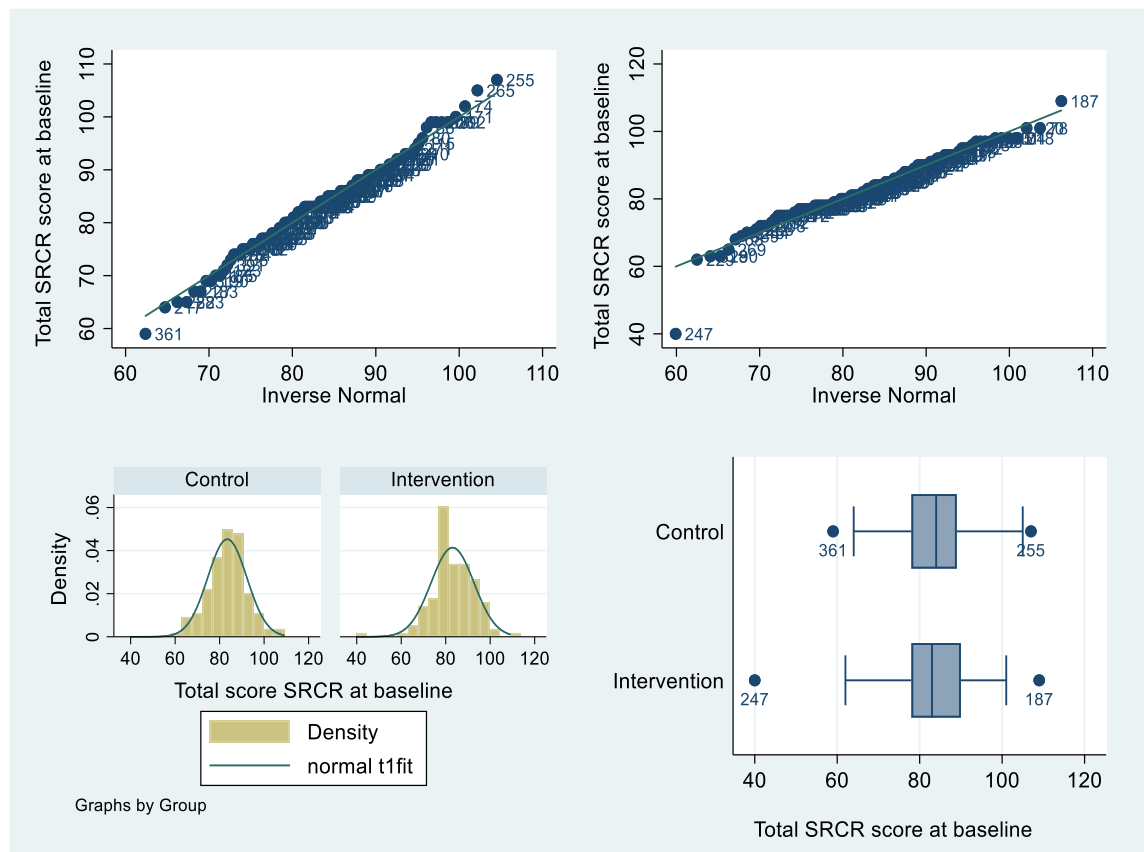
Note: SRCR refers to self-reported clinical reasoning skills

Figure 6-4 Q-Q plot, histogram and boxplot showing the normality of the distribution of FIT scores at Time 1 by group allocation
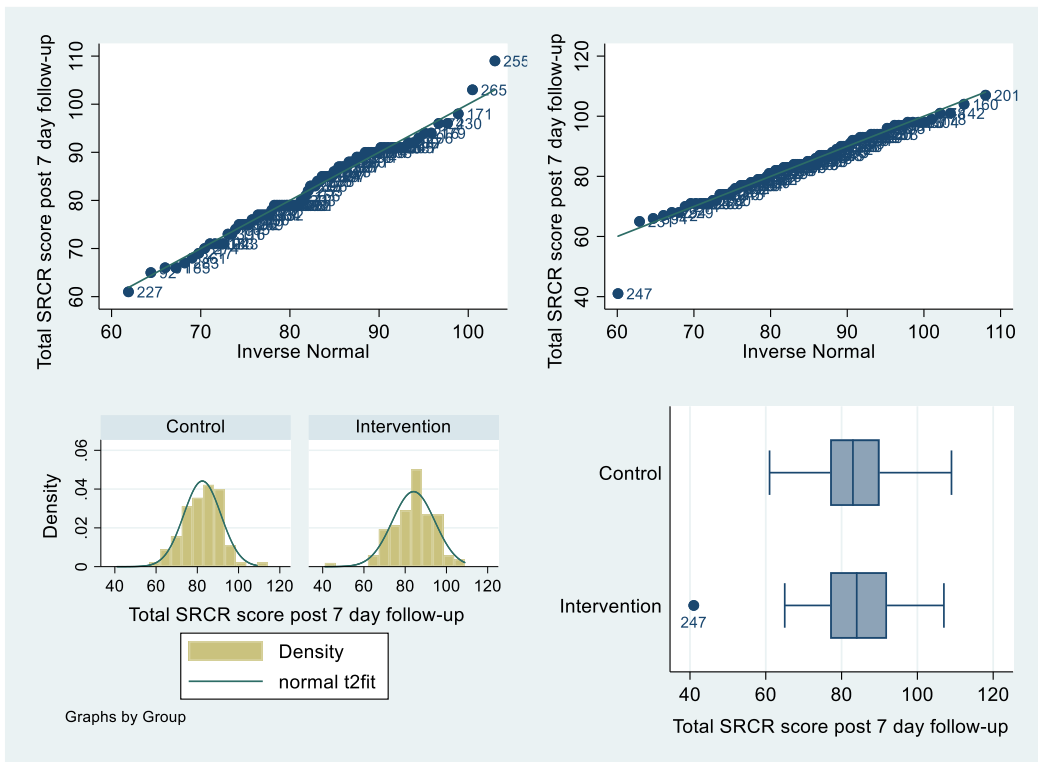
Note: SRCR refers to self-reported clinical reasoning skills

Figure 6-5 Q-Q plot, histogram and boxplot showing the normality of the distribution of FIT scores at Time 2 by group allocation
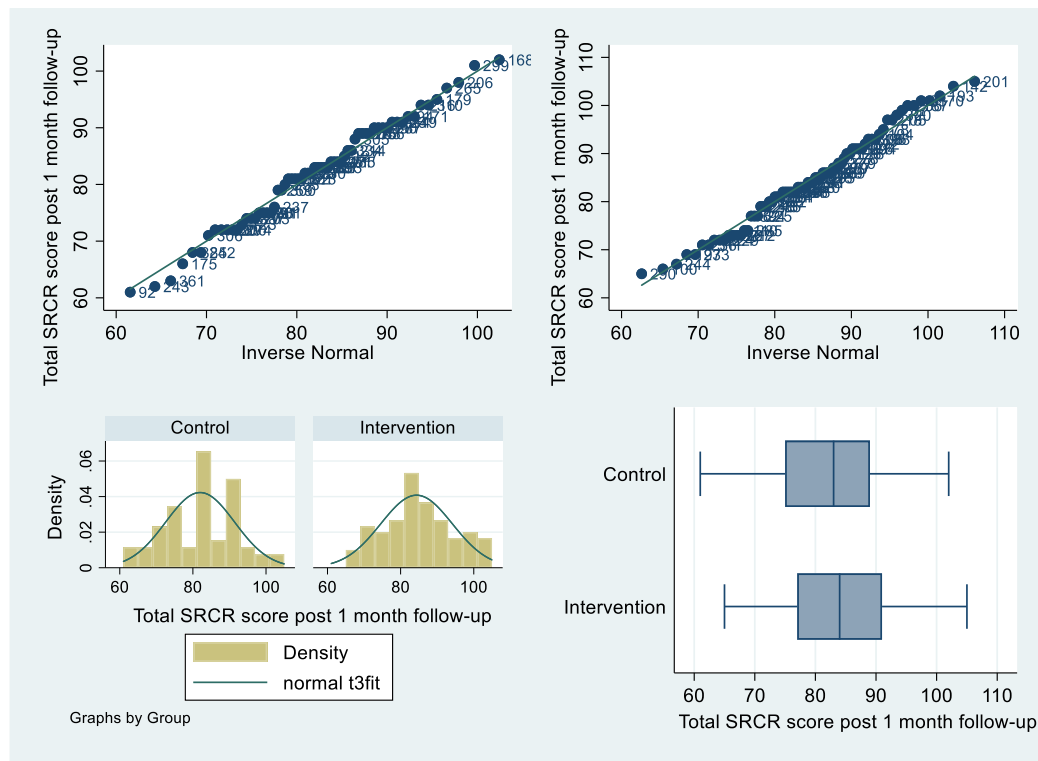
Note: SRCR refers to self-reported clinical reasoning skills

6.4.1.2.2 Homogeneity of variance and sphericity

The Levene's test was not significant for self-reported clinical reasoning skills at any time point, indicating that the assumption of homogeneity of variance was satisfied. The assumption of sphericity was also satisfied, as the estimates of sphericity were close to 1 (Huynh-Feldt epsilon = 0.98, Greenhouse-Geisser epsilon = 0.97).

6.4.1.3 Mixed factorial ANOVA comparing self-reported clinical reasoning skills over time and between the intervention and control group

A mixed factorial ANOVA was conducted to explore the main effect of time and group allocation on self-reported clinical reasoning skills, and the interaction between time and group allocation. A total of 136 students had data at all 3 time points when self-reported clinical reasoning skills were assessed (control $n = 62$, intervention $n = 74$). There was not a significant main effect of group allocation ($F_{(1)} = 0.00$, $p = 0.97$) or time ($F_{(2)} = 0.01$, $p = 0.99$) on self-reported clinical reasoning skills. There was also no significant interaction between group allocation and time, $F_{(2)} = 0.48$, $p = 0.62$. The intervention group appeared to score consistently higher than the control group on self-reported clinical reasoning skills. The estimated marginal means plot shows the average self-reported clinical reasoning skills between the intervention and control group over time (**Figure 6-6**). A sensitivity analysis was carried out by removing the four outliers identified in **Section 6.4.1** and re-running the analysis. The outliers had no effect on the results of the mixed-factorial ANOVA, so they were not removed from the analysis.

Figure 6-6 Predicted mean self-reported clinical reasoning scores over time and by group allocation

Predictive Margins of group#time with 95% CIs
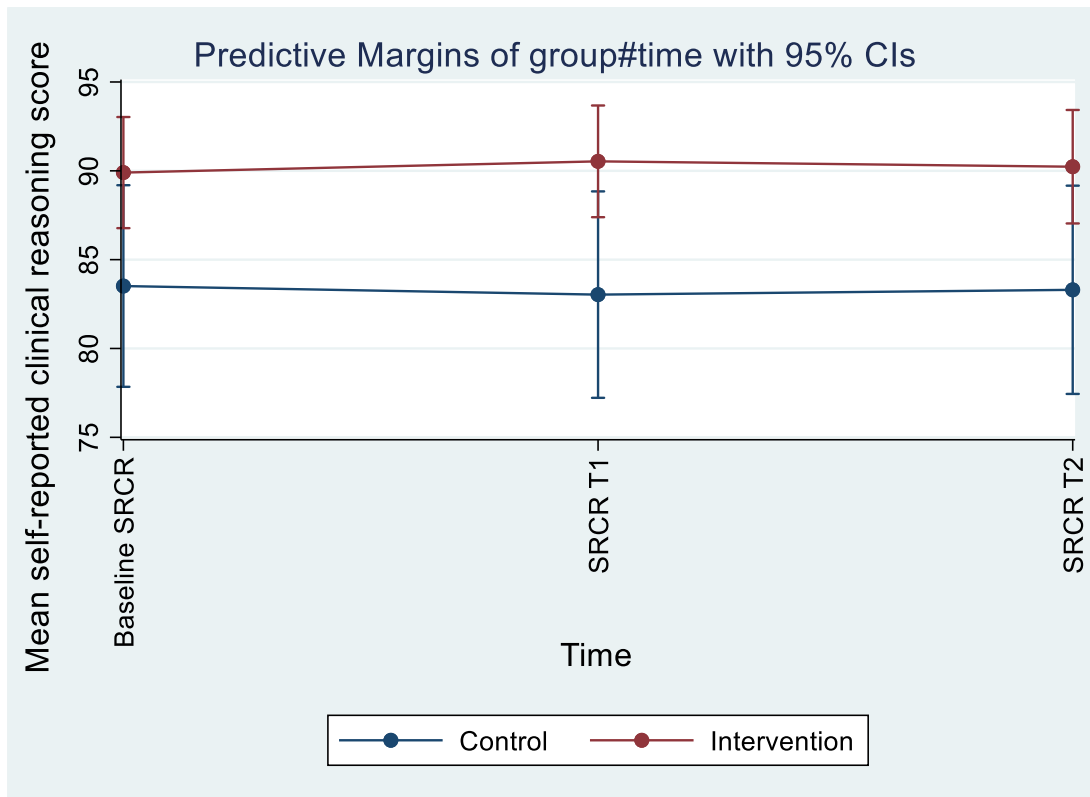
6.4.2 Observed clinical reasoning skills

6.4.2.1 Descriptive statistics for key features for patient case 4

Descriptive data showing the number of students and mean scores for the key features of patient case 4, by group allocation, are shown in **Table 6-13.** On all key features the intervention group scored higher than the control group.

Table 6-13 Descriptive statistics for key features of patient case 4 by group allocation

	Patient case	Trial group	*n*	Mean (*SD*)	Median (*IQR*)
Key Features					
1. Relevant information gathered (%)	4	Intervention	78	81.4 (10.5)	83.0 (15.1)
	4	Control	70	84.6 (10.6)	85.4 (13.9)
2. Essential information identified (%)	4	Intervention	78	61.6 (17.6)	62.1 (24.1)
	4	Control	70	53.3 (15.8)	51.7 (24.1)
3. Changed diagnosis (No. of times)	4	Intervention	78	3.2 (1.0)	3 (2)
	4	Control	70	3.0 (1.0)	3 (2)

6.4.2.2 Logistic regression analyses comparing the intervention and control group on the key features of case 4

The analysis for this section was completed for participants who had completed case 4 (*n* = 148) and were either in the intervention or control group. The outcome variables for the three regression models were the three key features. The only predictor variable inputted into each model was group allocation, as there were no significant differences between groups in baseline characteristics. **Table 6-14** summaries the beta coefficients, standard errors, odds ratios, 95% confidence intervals and *p* values for each regression model.

Table 6-14 Logistic regression analyses[1] comparing key features between the intervention and control groups

Key Feature		B (*SEM*)[2]	Odds Ratio (95% CI)	P value
1. Relevant information gathered		-0.23 (0.12)	0.79 (0.62, 1.01)	0.06
2. Essential information identified		0.34 (0.11)	1.40 (1.12, 1.75)	≤.01*
3. Changed	2 (base)			
	3	0.39 (0.40)	1.48 (0.68, 3.24)	0.32
diagnosis	4	0.49 (0.45)	1.63 (0.68, 3.92)	0.27
	5	0.90 (0.76)	2.46 (0.55, 11.00)	0.24
	6	0.21 (1.44)	1.77 (0.07, 20.76)	0.89

[1] In all models 1 = control 2 = intervention. $n = 148$
[2] For Key Features 1 and 2, standard errors reported are robust standard errors
* *P* value significant

The logistic regression analysis found there was no association between group allocation and how much relevant information students gathered for patient 4, but this association approached significance (χ^2 (1) = 3.44, p = 0.06). The predicted marginal means showed that the control group on average asked 4% more relevant questions, out of all the questions they asked for, than those in the intervention group (**Figure 6-7**). Studentised residuals revealed that there were no outliers that were likely to have influenced the analysis.

There was a significant association between group allocation and the proportion of essential information identified, out of all possible essential information for the case (χ^2 (1) = 9.0, $p \leq 0.01$). The odds of identifying a higher proportion of essential information from the patient increased by a factor of 0.40 for students in the intervention group versus the control group. The predicted

marginal means showed that those in the intervention group identified on average 9% more essential information than the control group (**Figure 6-7**). Studentised residuals revealed that there were no outliers that were likely to have influenced the analysis.

Figure 6-7 Predicted mean proportion for key features 1 and 2 by group allocation
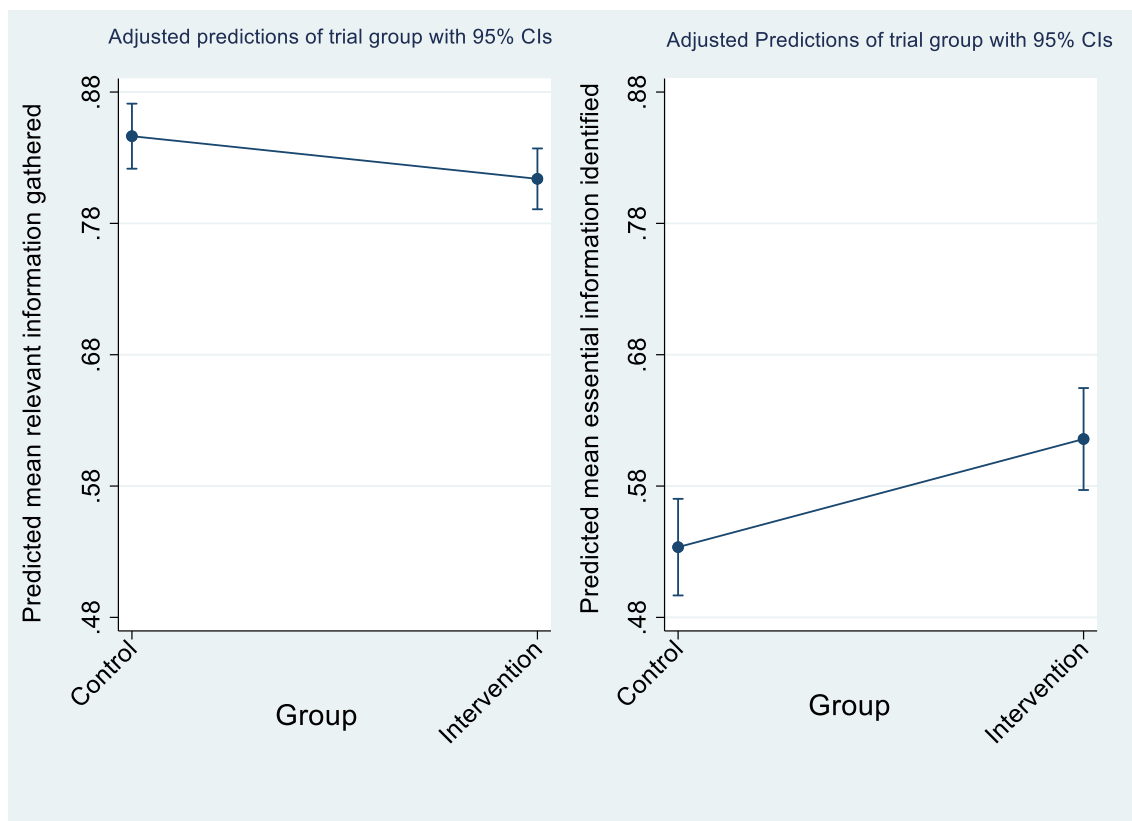
Further exploration of the data, using Mann-Whitney tests, revealed that the actual number of relevant questions and tests students in the intervention group gathered (*Mdn* = 21.5, *IQR* = 10) was significantly higher than those in the control group (*Mdn* = 18, *IQR* = 9), *z* = -2.72, *p* ≤ 0.01, *n* = 148. Equally, the intervention group also identified more essential items (*Mdn* = 18, *IQR* = 7) than

the control group (*Mdn* = 15.5, *IQR* = 7), *z* = -2.82, *p* ≤ 0.01, *n* = 148. The intervention group (*Mdn* = 25.5, *IQR* = 15) asked significantly more questions than the control group overall (*Mdn* = 20, *IQR* = 12), *z* = -2.83, *p* ≤ 0.01, *n* = 148 and asked significantly more irrelevant questions from the patient (*Mdn* = 4, *IQR* = 6) than the control group (*Mdn* = 3, *IQR* = 4), *z* = -2.52, *p* ≤ 0.05, *n* = 148.

Students in both groups changed their diagnoses at least twice. A multinomial logistic regression revealed that there was no significant difference in the number of times students changed their diagnoses between groups (χ^2 (4) = 2.24, *p* = 0.69). The predicted probabilities of the number of times students changed their diagnosis are presented in **Figure 6-8.** The predicted probability of making just 2 changes in diagnosis ideas was (non-significantly) higher in the control group (40%) than the intervention group (30%). The probability of changing diagnosis more 3-6 times was similar between groups; for example, 20% of controls changed diagnosis 4 times vs 25% in intervention group.

Figure 6-8 Predicted probability of key feature 3 by group allocation

Adjusted predictions of trial group with 95% CIs

Legend:
- Changed diagnosis twice
- Changed diagnosis 3 times
- Changed diagnosis 4 times
- Changed diagnosis 5 times
- Changed diagnosis 6 times
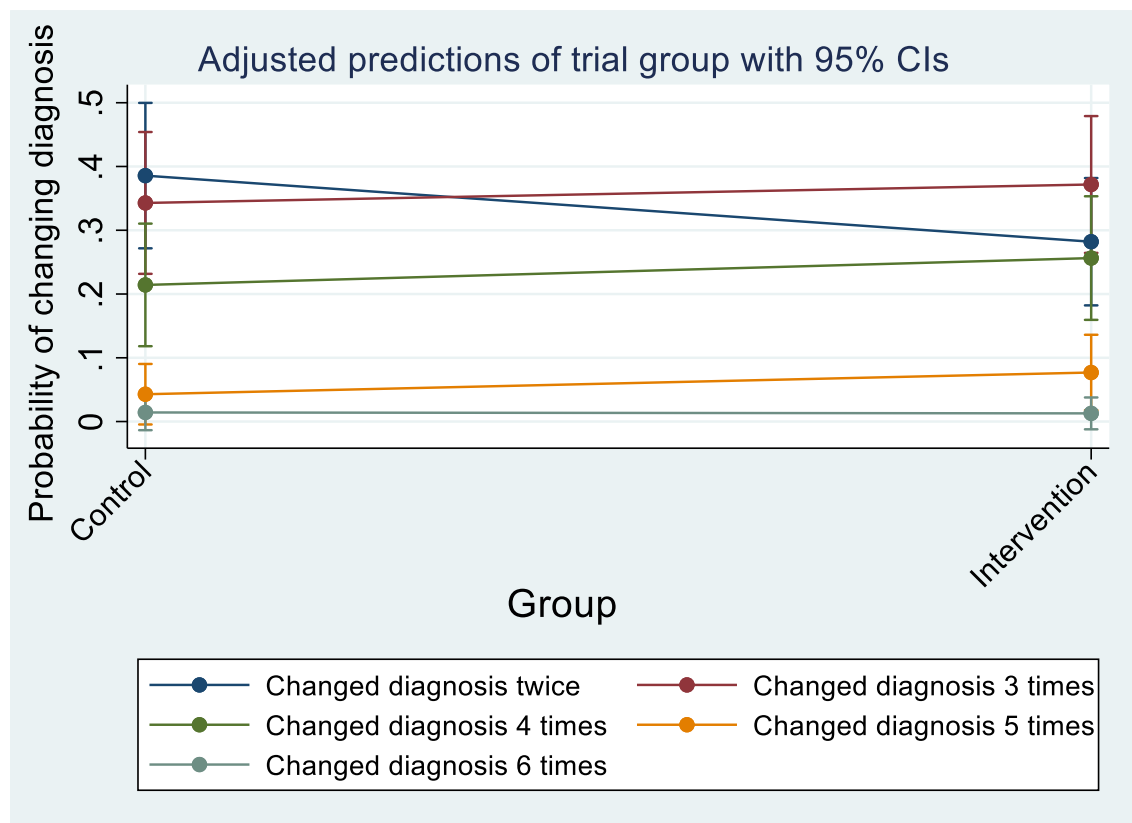
6.4.3   Observed consultation outcomes

6.4.3.1 Descriptive statistics for the consultation outcomes for patient case 4

Descriptive data showing the number of students and mean scores for consultation outcomes between the intervention and control group are shown in **Table 6-15.** The intervention group scored higher than the control group on most consultation outcomes, excluding the selection of appropriate further investigations and follow-up tests.

Table 6-15 Descriptive data for the consultation outcomes for patient case 4 by group allocation

	Trial group	*n*	Mean (*SD*)	Median (*IQR*)
Consultation outcomes				
1. Improved number of appropriate diagnoses	Intervention	23 (29.5)		
	Control	11 (15.7)		
2. Selected lung cancer in their initial diagnosis	Intervention	61 (78.2)		
	Control	54 (77.1)		
3. Selected lung cancer in their final diagnosis	Intervention	62 (79.5)		
	Control	52 (74.3)		
4. Percentage of final differential diagnoses that were appropriate	Intervention		82.7 (12.3)	83.3 (0)
	Control		82.4 (10.2)	83.3 (0)
5. Percentage of investigations that were appropriate	Intervention		78.6 (24.6)	83.3 (33.3)
	Control		82.4 (22.5)	100 (33.3)
6. Selected appropriate follow-up option.	Intervention	42 (53.9)		
	Control	40 (57.1)		

6.4.3.2 Logistic regression analyses comparing the intervention and control group on the consultation outcomes of case 4

The dependent variables for the six regression models were the six consultation outcomes. The only predictor variable inputted into each model was group allocation. All regressions were binary logistic regressions except consultation outcome 1, which was a multinomial logistic regression. **Table 6-16** summaries the beta coefficients, standard errors, odds ratios, 95% confidence

intervals and *p* values for each regression model. There was no significant association between any of the consultation outcomes, between the intervention and control groups.

Table 6-16 Logistic regression analyses comparing consultation outcomes between the intervention and control groups

Consultation outcomes		B (*SE*)	Odds Ratio (95% CI)	P value
1. Improved number of appropriate diagnoses	Less appropriate diagnoses at final vs initial (base)			
	Same number of appropriate diagnoses at final vs initial	-0.07 (0.84)	0.93 (0.18, 4.81)	0.93
	More appropriate diagnoses at final vs initial	0.74 (0.90)	2.09 (0.36, 12.08)	0.41
2. Selected lung cancer in their initial diagnosis		0.06 (0.40)	1.06 (0.49, 2.31)	0.88
3. Selected lung cancer in their final diagnosis		0.29 (0.39)	1.34 (0.62, 2.89)	0.45
4. Proportion of final differential diagnoses that were appropriate		0.04 (0.13)	1.04 (0.81, 1.34)	0.78
5. Proportion of investigations that were appropriate		-0.24 (0.25)	0.79 (0.48, 1.28)	0.33
6. Selected appropriate follow-up option		-0.13 (0.33)	0.88 (0.46, 1.68)	0.69

6.5 Discussion

6.5.1 Feasibility

The results of the feasibility trial indicated that it would be feasible to conduct a larger definitive trial to assess the effectiveness of eCREST on medical students' clinical reasoning skills in multiple medical schools. A full-scale definitive RCT would need 128 students, with 64 students in each group, to detect an 8.3% difference in mean proportion of essential information identified (key feature 2). This assumes $\alpha = 0.05$, 80% power to detect a true effect and a standard deviation of 16.7 (average standard deviation across control and intervention group calculated from this study). To account for the 30% of students that would be likely to drop out of the study after one-week, as found in this study, 183 students would need to be recruited to ensure that there would be 128 complete cases in the analysis. If the follow-up were to occur after one month there may be up to 50% of students who drop out of the study and, therefore, 256 students would need to be recruited.

I found students were willing to be recruited but successful recruitment depended on several factors. A key driver for successful recruitment was the time of year when students were recruited. Students in the autumn and winter terms were more likely to take part in and stay the trial than those in the summer. This was most likely because they found eCREST to be a useful resource for revision. Indeed, students who took part in the autumn and winter terms also perceived that eCREST had greater learning benefits than those who took part in spring and summer. Students who had completed their final examinations, and were about to go on their elective, may have had less motivation to take

part. While some may have viewed eCREST as preparation for practice, this might have been less motivating than preparation for exams. SB had the most successful recruitment, especially in cohort 2, and this was most likely due to the efforts of the tutor in promoting the study via Moodle and the student representative promoting it effectively. SA's recruitment was also more successful in cohort 2, which may be largely attributable to the incorporation of eCREST into a relevant module. It was unclear why students at SC had such low recruitment in comparison to the other sites when students were recruited in the same way as SB. A possibility is that students at SC may receive more patient cases than SB and SA as part of their PBL designed curriculum, so eCREST may have appeared less novel or useful. SA students may also have more actual patient contact because of course design and ease of patient access, reducing the need for simulated cases.

I found that around 20% of eligible students volunteered to take part in my feasibility RCT. Most previous studies I identified in my systematic review had greater uptake but had integrated the interventions into the curriculum, so completion was compulsory for students. In the few studies in my systematic review where completion was not compulsory the uptake was considerably lower. Lehmann et al. (2015) only achieved 11.9% uptake across two year groups but across one year group Wu et al. (2014) achieved 58% uptake and Devitt and Palmer (1998) achieved 85%. There was little other evidence of what expected uptake would be for voluntary use of an online patient simulation tool. However, evidence from studies of MOOCs found that uptake is much poorer than 20%, with around only 7% of students enrolling in the modules (Aboshady

et al., 2015; Freitas, Morgan, & Gibson, 2015). Uptake in my study may have been better than seen in MOOCs and some previous studies because monetary incentives were used. However, some students commented in their feedback that they would use eCREST again without an incentive and most stated that they used eCREST to help them prepare for clinical practice.

Unlike other studies of adult online learning that found around 40-80% of students drop out of online courses, I found there was a relatively low drop out from eCREST (Bawa, 2016). Only around 30% of students dropped out after one week and 50% dropped out after one month. Furthermore, there were no significant differences in retention between the intervention and control group, suggesting that students were engaged with the intervention. Indeed, most students completed the online cases they were assigned. However, it might be expected that uptake and retention rates would considerably improve if the intervention was integrated into the curriculum and consequently mandatory for students to complete. This would also have pedagogical benefits, as it would help students to understand how the teaching fits with their wider teaching and to join concepts together. Ideally, a definitive RCT would integrate online patient simulations into the curriculum to demonstrate effectiveness, within the context that it is intended to be used in.

6.5.2 Acceptability

My results concurred with findings from other studies that evaluated similar interventions with regards to acceptance of online patient simulations.

Students from previous studies also indicated that they felt online patient simulations are an acceptable way of teaching clinical reasoning skills and wanted to use more simulations in their curriculum (Aghili et al., 2012; Kleinert et al., 2015). Furthermore, wider research has shown that medical students value the use of eLearning in general to improve their clinical skills and view them as comparable to traditional ways of learning (Gormley, Collins, Boohan, Bickle, & Stevenson, 2009; Poulton, Conradi, Kavia, Round, & Hilton, 2009). Further to this, I found students thought eCREST was user-friendly, with almost 100% reporting that it was easy to navigate through and achieved the right level of difficulty. I also found that over 80% of students perceived eCREST to have helped them learn clinical reasoning applicable to their clinical work, as it helped them to structure their reasoning. However, there were constructive suggestions of ways in which eCREST could be improved, such as allowing students to have more input by allowing them to create their own diagnosis and question lists before seeing the options available in eCREST. These will be explored through consultation with students and medical educators for future iterations of eCREST.

It is important to acknowledge the increasing role that technology is playing in medical education and the potential benefits to students' learning and teaching, such as reaching a wider audience, adaptability and potentially lower costs for faculty in the long-term (Topol, 2018). However, the results from this study highlight some of the limitations of eLearning, in that it cannot fully replicate the learning that students need, which is face-to-face contact with real patients. The current level of sophistication of online patient simulations like

eCREST mean that the virtual interactions can be restrictive and not allow for students to think for themselves and provide more input. They also do not necessarily allow for much personalised feedback, which students need for reflection and learning. The findings from this study can provide broader recommendations for the teaching of clinical reasoning skills in medical schools, using not only virtual methods but a variety of learning methodologies. For example, senior medical students find it useful to be provided with the opportunity to reason and complete a task before receiving explicit guidance and feedback. They also prefer feedback to be more explicit and focus on the process of reasoning, such as highlighting points in which the student exhibits cognitive biases.

6.5.3 Evaluation of clinical reasoning skills

Similar to previous studies that have used the FIT or DTI survey to assess reasoning I found there was no significant difference in self-reported clinical reasoning skills over time or between groups (Bateman, 2013; Lee et al., 2010) but the intervention group in my study did appear to score consistently slightly higher scores than the control group. This could imply that the FIT survey is not a sensitive enough measure to detect what would probably amount to small changes in clinical reasoning skills over time. Additionally, it may be too difficult for students to think objectively about their clinical reasoning using self-reported measures retrospectively. Retrospective accounts of thinking are more likely to be subjective and confabulated because the information is influenced by other information in the long-term memory store (Ericsson & Simon, 1993). In my systematic review, I found that studies that evaluated online patient

simulations more commonly used observational outcome measures of reasoning, usually via observing students completing a patient case or key features problem. Indeed, my use of a self-reported measure could explain why my results differed from most of the studies identified in my review that compared online patient simulations to no formal instruction, which all found significant positive effects of the intervention between groups and over time (Kalet et al., 2007; Kleinert et al., 2015; Lehmann et al., 2015; Wu et al., 2014). More observational ways of measuring thought processes, such as measuring the performance on the key features of a patient case, may be more appropriate as they would assess thinking prospectively, which is likely to be a more accurate and realistic representation of reasoning.

Results from comparing student performance on the key features of patient case 4 revealed that students who were in the intervention group were significantly better at identifying the essential information needed to make an appropriate diagnosis than those in the control group. Interestingly, however, those in the control group appeared to gather more relevant information but the difference between the groups only approached significance. It seems the intervention group were better at identifying both relevant and essential information but they also tended to ask for more irrelevant information. This suggests that eCREST may help students to be more effective at identifying red flags or potentially serious symptoms that could be indicative of serious diseases. However, by doing this it may also make students less efficient in the way they gather information, by increasing the number of questions they ask. A potential reason for why the intervention group gathered data in this way could

be because the students that used eCREST before may have been expecting a complex and uncertain patient case, as they would have been primed based on the feedback of the previous cases on the need to consider potentially serious diseases even if they are unlikely. This may have changed the strategy of those who used eCREST before to be broader and more open-minded when gathering information, which may have led to more irrelevant information being gathered. Further discussion with medical education experts and clinicians is needed to fully understand whether this strategy is appropriate to develop for uncertain cases in primary care in real clinical practice for students at this level of education. Given the uncertainty prevalent in primary care it could be useful to approach cases more broadly and open-minded and, therefore, eCREST perhaps should continue to prompt students to ask more questions. It is also possible that teaching students to ask lots of questions is an effective learning strategy for students, at least initially. If students have more practice with additional patient cases this could help them to hone these skills and eventually become more efficient through repeated practice (Ericsson, 2008). Moreover, there may be additional features that could be added to eCREST to help students think about efficiency and reduce the amount of irrelevant information gathered.

Despite appearing to use different data gathering strategies, there was no significant difference between the intervention and control group in their consultation outcomes. I predicted in my logic model in **Chapter 4** that the intervention group should have better consultation outcomes than the control group because their data gathering should have improved as a result of

eCREST. Since this was a feasibility RCT it might not have been adequately powered enough to detect differences in consultation outcomes or it may be that no difference was found because eCREST only provides training on the processes of data gathering and not on training consultation outcomes like diagnostic accuracy and patient management. Perhaps other training in conjunction with eCREST is needed to improve students' knowledge of symptoms, diagnoses and management to also improve consultation outcomes. Alternatively, more feedback could be provided to students in eCREST on diagnoses and how to manage patients that present with symptoms indicative of potentially serious diagnoses, so they also improve their knowledge.

Students in the intervention group tended to change their diagnosis more than those in the control group but this relationship was not significant. Furthermore, those in the intervention group were non-significantly more likely to improve the number of recommended diagnoses from their initial differential diagnosis to their final differential diagnosis. This could fit with the assumption that intervention group students were being more open-minded when gathering data from the patient, then reviewing the evidence and changing their diagnoses accordingly. However, it might also suggest more could be done by eCREST to encourage students to be flexible with their diagnoses.

6.5.4 Strengths and limitations of this study

One of the main strengths of this thesis is that it presents the first feasibility RCT to be undertaken at multiple medical schools in the UK. Previous studies in my systematic review did not focus on feasibility. Effectiveness can be moderated by lack of engagement, so it is important to understand the engagement of medical schools and students to deliver and use online patient simulations to learn reasoning skills. Without engagement from faculty, to support and deliver online patient simulations, and from students, to use them as intended, it is not possible to understand their true effect on reasoning (Ellaway & Masters, 2008; Greenhalgh, 2001). The advantage of conducting the study across multiple sites was that it provided a greater understanding of how eCREST could be used by medical schools who follow different curricula and different styles of teaching clinical reasoning skills, thus, demonstrating generalisability of the intervention. A further strength of the RCT design was that students were randomised to the intervention or control group, which allowed for potential confounders, such as age and differing levels of knowledge or reasoning ability to be evenly distributed between the groups.

Nevertheless, there were some limitations of the feasibility RCT. The main limitation was that of selection bias. RCTs usually help to eliminate this bias but because my RCT could not randomly select students from the cohorts and relied on self-selecting student volunteers, it was biased towards students who were willing to take part. It can be assumed that students who self-selected to take part were more likely to be interested in research, eLearning, preparing for clinical practice, revision or a monetary incentive, than those who did not

elect to take part. It could be that had I randomly selected from the population, students' views of eCREST might be poorer than those in this study and eCREST may have had less of an effect on their reasoning. However, it was only possible to collect data in this way from medical schools, given that eCREST was a newly developed tool that had not been previously tested. I attempted to reduce further influence from selection bias by being blinded to group allocation. If a full-scale RCT is conducted it should aim to randomly select from cohorts or year groups to reduce the effects of this bias.

The study was also not designed to be adequately powered to be able to detect the effectiveness of the intervention. However, this is acceptable because feasibility studies are not used for this purpose but to test outcome measures to determine the sample size in a full-scale RCT (Arain et al., 2010; NIHR, 2016). Nevertheless, given the paucity of knowledge in this area identified in my review, research testing the effectiveness of eCREST would be of interest to educators and researchers.

The other limitation to using the feasibility RCT study design was that medical schools had reservations about using this method to evaluate eCREST. The major concerns were about the fairness of providing eCREST to only students in the intervention group. All students were provided with eCREST at the end of the study but there were some concerns it may unfairly advantage students who had not used it before exams. Thus, I ensured in the trial that eCREST was either used after exams (in cohort 1) or that the trial ended before exams (cohort 2). The ethical concerns around RCTs in medical education may

raise concerns over the willingness of medical schools to participate in a full-scale RCT. If eCREST were to be evaluated using a full-scale RCT, a crossover randomised design might have better uptake from medical faculty, as this design would involve exposing both groups to eCREST but at different periods (Sibbald & Roberts, 1998; Sullivan, 2011). The benefits to this research design over that of a parallel group RCT would be that each student would serve as their own control, removing inter-individual variability and requiring a smaller sample size. However, there is a risk that carryover effects would persist into the second assessment period and bias outcomes (Sibbald & Roberts, 1998).

A further limitation was the outcome measures I used to assess clinical reasoning skills. The Flexibility in Thinking (FIT) scale only measured one aspect of clinical reasoning skills from the Diagnostic Thinking Inventory (DTI) (Bordage et al., 1990). The other scale of the DTI measured the structure of knowledge in memory, which is also an important aspect of clinical reasoning skills. However, I chose to use only one scale to reduce survey fatigue based on feedback from users. I chose the scale related to thought processes, as eCREST was designed to target thought processes. It is also possible that completion of the survey before the study influenced students' clinical reasoning during eCREST, making them more likely to be flexible. The measures I developed to assess clinical reasoning skills (the key features) also had limitations. I demonstrated they had good face and content validity, as clinicians found the key features adequately represented clinical reasoning skills. Nevertheless, I failed to demonstrate that they had predictive validity and was unable to detect changes in performance over time.

The data analysis conducted with the key features may also have been limited by the use of a logistic regression to transform the proportion outcome data of the key features. This technique is not widely used in this field but was recommended by a statistician and wider literature and the assumptions of the models were met (Dixon, 2008; Warton & Hui, 2011). I was also not blinded to group allocation when I received the data, which could have biased my analysis of the data. I attempted to overcome this by following my data analysis plan and not seeking to analyse any spurious associations.

6.5.5 Summary

This chapter addressed my third research question by showing that it was feasible to conduct an RCT comparing an online patient simulation to no formal instruction, across three UK medical schools. Furthermore, I found that online patient simulations were acceptable to medical students and that they had some impact on their clinical reasoning skills, particularly their ability to identify essential information in a patient case. This evidence provides proof of concept for the use of online patient simulations in medical curricula but a further trial is needed to determine their effectiveness when they are integrated into the curricula. In a full-scale RCT, assuming similar uptake and completion to the feasibility RCT, 256 students would be needed to be recruited to detect an 8.3% difference in the average proportion of essential information identified after one month with $p < 0.05$ and 80% power of detecting a true effect. However, fewer students would be needed if a randomised crossover study design was used (Sibbald & Roberts, 1998). This study does not provide the

necessary evidence to address my other research questions relating to online patient simulations, such as how students reason when using an online patient simulation and how other factors affect reasoning. These research questions will be addressed in **Chapters 7 & 8** of this thesis.

7 Methods: Think Aloud and interview study

In this chapter I present how I addressed one of my thesis aims, which was to evaluate an evidence and theory based training to improve medical students' clinical reasoning skills, by answering my fourth and fifth research questions of how simulations can improve students' reasoning and the factors that influence this.

7.1 Background

The theoretical model that I outlined in **Chapter 1** described how students might use their clinical reasoning skills. The OPS logic model (in **Chapter 4**) described how an online patient simulation tool, such as eCREST, could help students to learn these skills and how this learning could be observed. Data from my feasibility RCT has helped me to explore how I can add to my theoretical model of how students reason by showing that students can use different strategies of gathering data, as those who used eCREST appeared to as more questions than those who did not. However, the data presented so far could not provide further insight into students' thinking, such as their rationale for asking questions or selecting diagnoses, or the factors that affected their reasoning. Thus, a further research method was needed to help explore how students reasoned when using eCREST in real-time and to complement previous data.

An ideal method to explore clinical reasoning skills was an in-depth qualitative study using the Think Aloud approach, in which students are

observed verbalising their thoughts in real-time while completing a task. This method gives direct insight into the clinical reasoning of medical students, as it provides access to their thought processes during a clinical case presented in eCREST (Ericsson & Simon, 1980). Gathering data using this method is potentially less biased compared to gathering interview data, as the Think Aloud approach captures students' immediate reactions rather than their retrospective thoughts or perceptions of clinical reasoning skills.

Qualitative data could also provide information on the factors that may influence how students reason and particularly their style of gathering data. The quantitative data suggested that eCREST might influence reasoning, as it may encourage them to ask more questions but also identify the most essential questions to ask. Students verbal reports of why they select questions and diagnoses in eCREST could inform the rationale for taking this approach or other approaches in eCREST. The design of eCREST, in particular the continuous prompts to reflect on differential diagnosis and the feedback, could influence their style of gathering data. Learning theories have also suggested that other factors relating to students' characteristics, such as their habits, confidence, emotions, attitudes and motivation to engage with the task, are also likely to influence thought processes and subsequent behaviour, which could be detected via students verbalisations in the Think Aloud (Azevedo, 2005; Zimmerman, 2008).

7.2 Aims and objectives of the Think Aloud and interview study

One of the aims of my thesis was to evaluate a training tool based on evidence and theory to improve the clinical reasoning skills of medical students. To achieve this aim, I needed to answer my fourth and fifth research questions, which were to explore how students reason when using an online patient simulation and how other factors, such as the design of the online patient simulation and confidence in skills, could affect reasoning. Therefore, the objectives of this Think Aloud an interview study were to:

a) Collect and analyse qualitative data to identify key themes across the data and the factors that may influence reasoning, such as the design of eCREST, and personal factors, such as confidence and motivation;

b) Identify different styles of data gathering based on the quantitative data collected by eCREST on students' data gathering;

c) Organise themes by the different styles of data gathering to further define and explore differences between the styles.

7.3 Mixed-methods approach

I used a mixed-method approach to this study, as I was guided by my analysis of the quantitative data from the feasibility trial (**Chapters 6 & 7**), in the analysis and interpretation of the data from the qualitative study. Therefore, I used a sequential explanatory design (Creswell, Plano Clark, Gutmann, & Hanson, 2003; Johnson & Onwuegbuzie, 2004). The process was not entirely

sequential, as the qualitative data was collected before the quantitative data was analysed; but the qualitative data was only analysed after the quantitative data was analysed. **Figure 7-1** shows a diagram that outlines the sequential explanatory design in the context of this thesis. There was not equal emphasis on the two types of data, as the initial aims of the thesis had a quantitative orientation and the qualitative data was used to help explain the findings of quantitative data and further the theoretical model of clinical reasoning and OPS logic model.

Figure 7-1 Sequential explanatory design
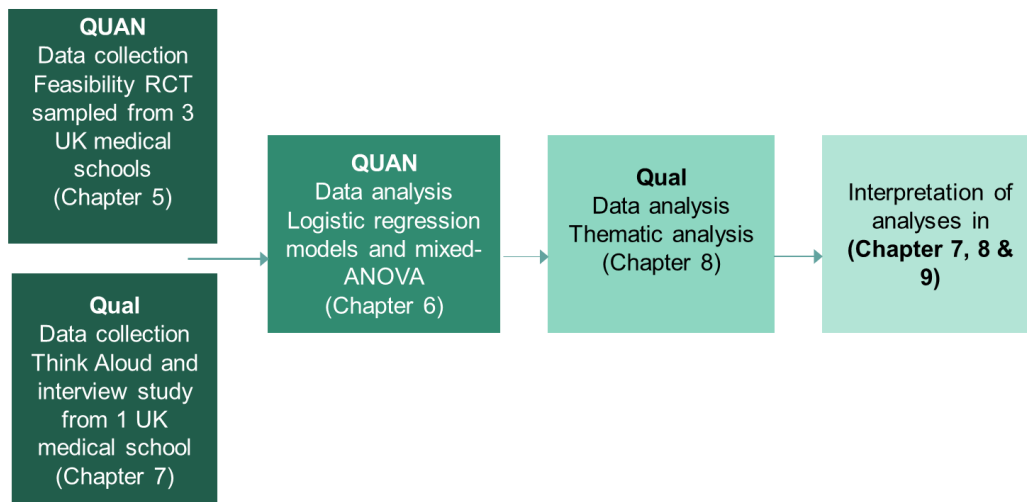
7.4 Study design

I selected a Think Aloud study design followed by a semi-structured interview study (Ericsson & Simon, 1980; Ericsson & Simon, 1993). Participants can be prompted to speak in the Think Aloud if they are silent for a significant amount of time but the prompts must be non-directive, such as 'keep talking',

to ensure attention of the participants is not redirected (Ericsson & Simon, 1993). Semi-structured interviews are often used after the Think Aloud to follow-up on comments made in the Think Aloud and to explore students' retrospective thoughts on their performance on the task (Ericsson & Simon, 1980; Ericsson & Simon, 1993). I used a semi-structured interview after each Think Aloud for this purpose.

Previous research has used the Think Aloud approach to explore clinical reasoning skills of health professionals and students. Studies comparing the reasoning skills of different health professionals found that both nurses and general practitioners make decisions from a pattern matching process, as a result of previous experience and education. Furthermore, they found there was little difference in clinical reasoning skills between nurses and general practitioners in scenarios where both practitioners had equal experience of the problem at hand (Offredy, 2002; Offredy & Meerabeau, 2005). Other research has investigated the differences between doctors and medical students in clinical reasoning and focused on how clinicians use different lines of reasoning to students (Johnson et al., 1981; Johnson, Hassebrock, Duran, & Moller, 1982). They found that students have more limited lines of reasoning when faced with unusual patient data because of their limited knowledge of diseases. They tend to take a more data-driven approach to problem solving but often miss unusual patient data, so their solutions are often incomplete. Experts tend to notice unusual patient data and use their global knowledge of diseases and the patient to make decisions, resulting in less missed diagnostic opportunities (Johnson et al., 1981; Johnson et al., 1982). Previous studies were not carried

out using online patient simulations, so the reasoning of students during these types of clinical cases is still unknown.

7.5 Sampling and recruitment

I used a mixture of convenience and snowballing sampling to recruit final year medical students at SA. I initially recruited PALs students at SA medical school. This scheme is an optional four-week module that allows students in their final year of study at SA to undertake a project in medical education. I advertised for three PALs medical students to: be participants in my Think Aloud study; assist in the recruitment of other final year students and help conduct the Think Aloud and interview studies. I offered them the experience of recruiting and conducting the Think Aloud to provide students with experience of recruiting and conducting research in medical education, which fulfilled the requirements of PALs. Three PALs students were recruited, and each student recruited two further students to take part in the study before their module was completed. Additionally, I asked if the PALs students could pass on details of the study to any of their peers who may be interested, so I could continue to recruit students after their module ended - this resulted in one more student being recruited. I also recruited students, after the PALs module had ended, via The Royal Free, University College and Middlesex Medical Students' Association (RUMS) news bulletin. In summary, three of the participants were recruited from PALs, seven students were recruited by the PALs students, and six were recruited via the RUMS bulletin.

All participants, except the three PALs students, received a £45 Amazon gift voucher for taking part in the study. The use of incentives may have affected students' decisions to take part in the study but were deemed necessary given the study would require the students to travel to a SA campus and give up at least one and a half hours of their time to be observed and recorded thinking aloud. Furthermore, I was interested in capturing the views of those whose only motive to take part was the monetary incentive. Students who were motivated by the monetary incentive could represent students who have low engagement with teaching or eLearning. Given that there will always be a proportion of students who do not engage in some types of teaching, it would be interesting to explore the reasons for this so adaptations to teaching could be made.

To determine the sample size I looked at previous Think Aloud studies in the area to get a sense of when theoretical saturation is likely to occur in similar studies; previous studies recruited approximately 12-21 participants (Johnson et al., 1981; Johnson et al., 1982; Offredy, 2002). I set out to recruit at least 16 participants, so that at least three different medical students were observed completing the four patient cases available in eCREST. I wanted to observe students completing all the different patient cases to capture the variety of clinical reasoning skills that students might use during eCREST, which could differ depending on the clinical case.

7.6 Planning

I developed a topic guide for the Think Aloud and semi-structured interview (see **Appendix 9**). The topic guide with regards to the Think Aloud

study contained suggestions of prompts and an introduction and practice exercise to the Think Aloud method, to ensure that participants understood the method. The topic guide for the semi-structured interviews contained questions to ask the participants. Questions were developed to address the research questions by trying to explore how they felt eCREST helped or hindered their reasoning and how they usually reason. I picked up on points that they seemed to find difficult in eCREST during the Think Aloud to allow them to explain their reasoning and why it was difficult. To help inform the OPS logic model I also asked questions about the personal factors that may have influenced their clinical reasoning skills, such as their confidence to complete the task, the difficulty of the task, their motivation to use eCREST again, and how realistic they found the simulation.

I piloted the topic guide in **Appendix 9** and the Think Aloud and interview with a GP registrar (SM) who was working in the research team. I did this to practise the Think Aloud approach, test the usefulness of the topic guide and test the acceptability of the questions I would ask in the interview. I asked for feedback on how I conducted the Think Aloud and interview, in particular how they felt about my prompting them to think aloud and the questions I asked. I also piloted the study with two SA medical students in their fifth year of study, to see whether medical students would be willing to think aloud and would not be inhibited by the scenario. Additionally, I piloted the Think Aloud to gain insight into the quality of data that would likely result from the Think Aloud when used with eCREST. I amended my topic guide based on the responses and recommendations of the pilot participants by making some of the questions

clearer and breaking down some of the questions in the interview. I also changed the introduction to the Think Aloud study by including some more explicit examples of things students might say and directed them to talk about their clinical reasoning skills specifically. This was to make it clearer to the students what they had to do and to put them at ease, as some seemed unsure of what they needed to verbalise in the Think Aloud. I also learned I needed to be more active in prompting students who had become too silent, as I found this challenging with some pilot participants.

7.7 Ethics

Ethical approval for this study was obtained from SA Research Ethics Committee via an amendment to the ethics application for the feasibility RCT described in **Chapter 5** (Ref: 9605/001; 8th September 2017). There were some ethical issues raised by this study, as students were being recorded and observed while using eCREST. To ensure student confidentiality and anonymity, only anonymised quotes were used in the data analyses and results and data were stored securely.

7.8 Procedure

Participants were required to register with eCREST by providing a username and password to create an account. They were instructed to select the option on the registration page for using eCREST as part of the Think Aloud study. They were then asked to complete a respiratory knowledge quiz and a clinical reasoning survey (FIT survey, see **Appendix 5 & 7**) (Bordage et al.,

1990). This was completed online and remotely, before the Think Aloud study. All students completed one patient case during the Think Aloud study. I randomly assigned a patient case to each participant before students took part. The patient cases were divided equally between the students so that at least four students did each case. Students practised the Think Aloud technique for approximately 5-10 minutes, on an unrelated web-based task (ordering a train ticket), which has been used in previous Think Aloud studies. This was to familiarise the student with thinking aloud, ensure they understood what to do and allow them to ask questions. If the student understood the procedure, they were provided with the eCREST website link (http://silverdistrict.uk/ecrest/), asked to login and started the study. They were instructed to complete their assigned patient case in eCREST and think aloud constantly while doing so. They were prompted if they become silent. In addition, they were asked some questions after they completed the task to explore comments they made whilst thinking aloud and what they thought of eCREST. The Think Aloud and interviews were audio recorded via Dictaphone and took approximately an hour and a half per student. For six of the participants, the PALs students assisted with this procedure.

7.9 Analysis

7.9.1 Thematic analysis

A professional transcription company transcribed the Think Aloud and interviews. I analysed the qualitative data from the Think Aloud and interviews thematically using a hybrid approach, involving deductive and inductive reasoning as described by Swain (2018). This approach follows similar

principles to that described by Braun and Clarke (2006), in that it describes the use of inductive and deductive coding, but is designed to be more flexible and explicitly describes how one may practically go through the process of coding and organising data using tables. I chose this approach as while I initially created a framework of deductive codes, based on the learning objectives of the tool and my logic model, I added inductive codes to these based on the data and codes identified by other coders. I found Swain's approach more practical and offered more explicit guidance for organising my codes and generating themes than other literature available. I followed the three phases and seven stages of analysis outlined by this approach and shown in **Figure 7-2** (Swain, 2018). I thematically analysed the data before defining data gathering strategies from the quantitative data to ensure my findings were not biased. In the fourth phase of analysis I created typologies based on students' quantitative data and then reorganised my current themes to compare these themes across typologies.

Figure 7-2 The three phases and seven stages of analysis of thematic analysis (Swain, 2018)

PHASE 1

Stage 1. Prepare a table/spreadsheet | Stage 2. Create a priori codes/themes | Stage 3. Begin to familiarise yourself with the data

PHASE 2

Stage 4. Begin a priori and a posteriori coding, creating a posteriori codes as you go along, and add them to the table/spreadsheet | Stage 5. Begin and then continue to add and summarise information from interview transcripts onto the table/spreadsheet

PHASE 3

Stage 6. Cut and paste text into a Microsoft WORD document | Stage 7. Collapse the a priori and a posteriori codes into family codes

Note: Image taken from Swain (2018)

7.9.1.1 Phase one

Firstly, I prepared a table that included the deductive codes based on my knowledge of the task, research aims and the expected mechanisms of change from my logic model. I split the task down into its component parts, focusing on the thought processes that were occurring at each step of eCREST. These steps were: generating hypotheses, gathering data, reviewing diagnoses, managing the patient, feedback and reflection. The reflection stage included the post Think Aloud interview, as students also gave reflective comments on their clinical reasoning and how eCREST influenced this in the interview. At each stage of the process, I created codes to identify any points where students explained their rationale for selecting questions from the patient, when they

commented on their performance and when they referred to eCREST explicitly. I then familiarised myself with the data to help develop more codes. I used NVivo 12 software to code all the transcripts.

7.9.1.2 Phase two

I coded according to my deductive codes that I had developed, and I created inductive codes based on patterns I began to see in the data as I coded each transcript. To ensure validity of the coding framework and that all codes were captured, three additional researchers (MK, APK and JT) read a different transcript each. Two of the researchers (JT and APK) used my initial coding framework to help them code the transcript and generated their own codes/themes. Importantly, JT is a GP registrar, so provided valuable insight from a clinical perspective, and was able to identify unique codes and provide greater insight into the students' knowledge. For example, JT suggested codes for the framework, such as implicit line of reasoning and wrong knowledge, which required medical knowledge to identify. Another researcher (MK) generated their own high-level codes/themes based on the transcript. Codes were cross validated in a meeting and I drew up a final coding framework based on the consensus in the discussion meeting. JT analysed a further transcript with the final coding framework to ensure that the data could be captured reliably, and no codes were missed. The framework was further validated in a meeting and minor amendments were made to the final coding framework based on their comments (see **Appendix 10**). I returned to earlier transcripts to ensure all inductive codes identified in later transcripts were captured.

7.9.1.3 Phase three

After reflecting on the codes across the phases of the task I collapsed the smaller codes into family codes or themes that spanned across each phase of eCREST. The grouping and interpretation of the themes was discussed in supervision panel meetings to refine the themes and I received feedback on the clarity of the themes. I used tables in Microsoft Word and Excel to organise and summarise my themes, as recommended by Swain. I identified six themes in the data, which are described in my results in **Chapter 8**. One of the key themes I identified in the data was that students were using different strategies for gathering data. This led me to explore in their quantitative data, collected as part of the task, how they gathered data. For example, I analysed their results for key features one to three, to explore how relevant their history taking was and how many questions they asked.

7.9.1.4 Phase four

In the final phase of analysis, I combined findings from qualitative and quantitative analysis. My findings from the qualitative data identified that students may have different approaches to gathering data. It seemed that there were variations between students in the relevance and focus of the information they gathered and the breadth of information they gathered, and that eCREST may have influenced these strategies. For example, some students were gathering relevant information but asking many unnecessary questions as well, whereas others were gathering relevant information but missing some essential information because they did not ask enough questions. To explore these

different strategies further I used the quantitative data from the task to categorise different ways students' reasoned during eCREST.

Creating typologies from quantitative data and exploring themes across these typologies is a common way of using a sequential explanatory mixed-methods design (Caracelli & Greene, 1993; Creswell et al., 2003). I created the typologies of different data gathering styles by analysing their data for key feature one, which measured how much relevant information the student asked the patient, out of the total number of questions they asked. I also measured how much essential information the student identified, out of the entire essential information available for that case (key feature two). Students' data gathering styles were categorised based on these key features because I found significant (or nearly significant) differences between the control and intervention group on these key features in my feasibility trial. I used the other relevant quantitative data, such as the key feature three and consultation outcomes, outlined in **Chapter 5**, to inform the themes I had identified through thematic analysis of the qualitative data from the Think Aloud and interviews. I then used the qualitative data to further describe and compare these different styles of reasoning and explored why students gathered information in different ways.

7.10 Reflections during fieldwork and analysis

Reflexivity in qualitative research is important to legitimise, validate and question the research process (Hennink, Hutter, & Bailey, 2010). I sought to be aware of the ways in which my sociocultural background, experiences, choice of research method and role as an observer may have influenced this research.

For the Think Aloud study, I was an observer and my presence was likely to affect the task. Students may have felt pressure to respond with socially desirable comments and to portray a certain level of knowledge to seem more able. I tried to attenuate possible social desirable responses by reassuring students in the information sheet, and when meeting them in person, the exploratory nature of the research. I reiterated the purpose was to explore how eCREST worked, rather than how well they can diagnose patients. I hoped this would allow students to feel more relaxed and less pressured to reach a 'correct' diagnosis in the task. After the task, I also reminded students of the GP feedback from the task. This stated that every clinician would approach the case differently and potentially have different diagnostic hypotheses. I also emphasised that it is difficult to complete a task while being observed and thinking aloud. My background as a non-medical professional/student, and being of a similar age to the participants, may have also helped to reduce socially desirable responses. My position meant that I had limited knowledge of what the most appropriate clinical decisions were, so I was not directly judging their clinical decisions. Students appeared open and relaxed during the Think Aloud and interview, which I interpreted as their feeling comfortable with my presence and position as a non-clinical researcher. This was also perhaps reflected in the fact that many students felt comfortable expressing knowledge and skills gaps that they could improve on (see **Chapter 8**).

Another possible impact I may have had on the research was that student responses might have been affected by their awareness that I was one of the developers of eCREST. Therefore, they may have been more positive towards eCREST than they would have done if unobserved or observed by someone external to the project. During the Think Aloud, they may have paid more attention and been more engaged more with eCREST, than they would have done if completing independently. In the interview, where I asked for student opinions of eCREST, they may have also not wished to offend me by expressing negative comments related to eCREST. To attenuate these socially desirable responses I emphasised to the students before the Think Aloud and interview that they should feel welcome to express negative comments or issues that they had with eCREST. I hoped this would allow students to speak openly about eCREST. Students responded with a variety of positive and negative comments about eCREST in the Think Aloud and interview, which I interpreted as their feeling able to express their opinions and engagement with eCREST. Participant five even felt comfortable expressing disinterest in the task (see **Chapter 8**).

There were alternative approaches to the Think Aloud that I could have used to capture thought processes that would have had different impacts on the data. One approach could have been to ask participants to provide an unprompted narrative account of their thought processes after completing the task (Adams et al., 2014; van den Haak, De Jong, & Jan Schellens, 2003). This approach may have minimised my impact on the participants, as students would not necessarily have needed to be observed, so less socially desirable

responses may have occurred. I would also have not needed to prompt them to think aloud during the task, therefore, I would have been less likely to guide their thought processes. Additionally, it would have reduced the cognitive load of the task, as thinking aloud adds another element to the task that could have interfered with the way students used eCREST. Retrospective accounts also provide the opportunity for students to reflect and provide more insight on their thought processes (van den Haak et al., 2003). However, there are benefits to using Think Aloud over retrospective accounts. Retrospective accounts only capture participants' retrospective thoughts, which can be influenced by other information in the long-term memory store (Ericsson & Simon, 1980; Ericsson & Simon, 1993). Additionally, students' retrospective accounts of using eCREST could have diverged from their thought processes during the task because they would have been influenced by receiving feedback. Another limitation of collecting retrospective accounts of thought processes is that participants can find it difficult to recall their thought processes during the task. I chose to use the Think Aloud method because it allowed me to access students' real-time thoughts while using eCREST. Additionally, I chose to interview students after the task to gather their reflections and further insights on their thought processes that could not be captured in the Think Aloud. I felt this approach enabled me to address my research questions. Nevertheless, the limitations of the method and the impacts of using this method on the data need to be explicitly acknowledged.

One of the most significant limitations of the Think Aloud approach is that it creates an artificial scenario in which participants are required to speak aloud

their thoughts while completing a task. Therefore, the method inherently influences students' thought processes and how they express these thoughts during the task. It is common in Think Aloud studies for observers to use generic prompts, such as "keep talking", to guide participants and ensure there are no silences. Often more directive prompts, such as "what did you think of that information?", are also necessary to remind participants to think aloud about the specific task (Charters, 2003; Cotton & Gresty, 2006). Generic prompts have been found to not significantly distort thought processes (Ericsson & Simon, 1980; Ericsson & Simon, 1993). However, it is debatable the extent to which more directive prompts can interrupt and influence the participant's attention and thought processes (Charters, 2003). Similar to others who have used the Think Aloud to explore online learning, I found while practising the method that students struggled to know what to think aloud and would often fall silent, particularly at the beginning of the task (Charters, 2003; Cotton & Gresty, 2006; Ericsson & Simon, 1980). This indicated to me that to obtain richer data about thought processes during the task I would need to actively prompt students, particularly at the beginning of the task. I used phrases such as "Why did you choose that?" and provided some examples in the instructions of what kind of thoughts they might speak aloud. I felt that the participants were more comfortable with the task when I provided these prompts. I attempted to attenuate the influence of the prompts in changing their thought processes by ensuring the prompts were open questions and as generic as possible. However, my approach likely meant that I did not fully capture some thought processes relevant to the task, as I influenced the thoughts they chose to speak aloud and interrupted the flow of their thoughts while doing the task.

Other approaches to observation have been used in Think Aloud studies that could have minimised the impact of the researcher on the data and provided other kinds of rich data related to thought processes. For example, some Think Aloud studies use video-recording rather than audio-recording. Video-recording can provide a richer and more complete data set than audio-recording alone. It reveals more data about non-verbal gestures, how the student is interacting with task and can be viewed by external observers to verify interpretations of the data (Bezemer et al., 2017; Bezemer & Mavers, 2011; Cotton & Gresty, 2006). As video-recording can capture non-verbal gestures, it potentially removes the need for a researcher to observe participants in the room to capture that data. On the other hand, it could be difficult to observe thought processes using the Think Aloud approach without having someone in the room prompting participants to think aloud. A further drawback of using video is that it is a time and resource intensive way to collect data, and the volume and multimodal nature of the data is complex to analyse (Bezemer & Mavers, 2011). I chose to use only audio-recording, as it was the most pragmatic approach given the timeframe and resources available for my PhD. However, I sought to note down non-verbal gestures and interactions with tool that I thought might not be captured by the audio-recording. On reflection, I also felt that video-recording the study would not necessarily have meaningfully added to the data. The simulation presented by eCREST does not require students to physically interact with a patient. Therefore, non-verbal information might have been less informative than it could be for other types of simulations where there may be physical interactions. Nevertheless, video-recording would

have provided a more complete picture of clinical reasoning skills when using eCREST. It may have been particularly informative for the additional coders in my study, allowing them to gain a richer understanding of the data and further verify my interpretation of the data.

At the end of almost all the interviews, I felt that students developed a good rapport with me in the interviews, as they all were happy to explain their thinking and discuss their training in the medical school. The only exception may have been with participant five, who was identified as a student who had a poor attitude towards the task and the patient case (see **Chapter 8**). They verbalised very little during the Think Aloud, even after prompting, and did not provide very detailed answers in the interview after the task. This person may have been more incentivised by the voucher to take part than their desire to test a new technology or develop their own learning. Overall, this person did provide a useful perspective of a disengaged student, which will reflect some students' use of eCREST if it were to be part of a medical education curriculum.

I also should acknowledge that I might have been influenced in the analysis of the qualitative data by my experience of analysing the quantitative data from the feasibility trial, which I did previously to analysing the Think Aloud and interview data. Indeed, this was part of my research design in that the quantitative analysis heavily influenced the analysis of the qualitative data. There may have been other meaningful ways of analysing the qualitative data, which would have provided different insights. However, this method was the most appropriate given my research questions. Furthermore, I initially analysed

the qualitative data thematically without referring to the quantitative data and generated inductive themes, to ensure I was not just noticing themes that I expected to be different across groups.

I found it challenging to describe the process by which I analysed the data and how this related to the logic model and its development. The steps described by (Swain, 2018) helped me to practically create the coding framework and add or change the codes based on the data. I took a sequential approach to the mixed-methods study, in that I analysed the quantitative data from the trial before analysing the qualitative data. However, in practice after analysing the qualitative data I also went back to the quantitative data collected in the Think Aloud task, to explore and create data gathering typologies that I had begun to see emerge in the qualitative data. These typologies were further explored and described in more detail with reference to the qualitative data. The qualitative data also influenced the interpretation of quantitative data in the trial (see **Figure 7-3**). Thus, the process described in **Figure 7-1** is not complex enough to describe the additional stages and interaction between the qualitative and quantitative data. The process was more iterative than sequential. The nature of the complexity of conducting mixed-method study made it challenging to present in a narrative in the thesis. This is problematic for demonstrating the rigour of the method used and for how useful my study is for future researchers looking to replicate or follow the process. However, the explicit description of the processes and an updated process diagram of how I conducted my mixed-methods study, as seen in **Figure 7-3**, could help to demonstrate rigour and help to inform other researchers of the mixed-methods process.

Figure 7-3 Updated process of mixed-methods approach
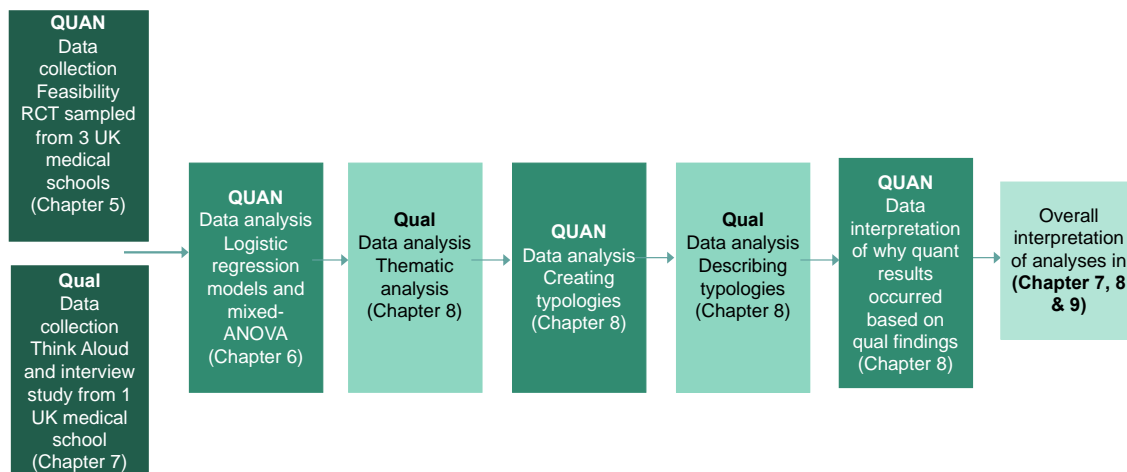
7.11 Summary

In this chapter, I discussed the rationale for using a qualitative research design to complement the quantitative data gathered in the feasibility RCT and to answer my fourth and fifth research questions pertaining to how students reason when using an online patient simulation and how other factors affect reasoning. I described how the different typologies of data gathering style were identified from the quantitative data and how themes identified in the qualitative data were compared between these typologies. I concluded by describing how I may have affected the collection and interpretation of the data and measures I took to ameliorate these effects. The next chapter (**Chapter 8**) describes the results of Think Aloud and semi-structured interviews.

8 Results: Think Aloud and interview study

In this section, I present the results of the Think Aloud and interview study. I first describe an overview of the participants and broadly outline the themes identified through thematic analysis. I then describe how the quantitative data informed the development of typologies for different data gathering strategies and then compare themes across these typologies. I also describe themes that relate to my fifth research question on how other factors affect reasoning and compare these across typologies. I conclude by summarising my typologies and referring back to my theoretical model of reasoning to explore what this study adds to the model.

8.1 Overview of participants

Between September and October 2017, I conducted 16 Think Aloud and interviews. For six of these Think Aloud and interviews a medical student from PALs assisted me (see **Chapter 7** for details). Slightly more males (n = 9/16) took part than females (n = 7). The average age was 24 years of age (SD = 1.1). The average baseline MCQ score was 9.9 (SD = 1.0) and average FIT score was 80.1 (SD = 6.7). The median length of the Think Aloud protocol and interview was 61 minutes 4 seconds (IQR = 19.63).

8.2 Thematic analysis

An overview of themes can be found in **Table 8-1**Error! Reference source not found.**.** Theoretical saturation was achieved after 16 participants took part, as

no new data emerged in any of my themes. There were six themes identified from the qualitative data that relate to the measurable outcomes identified in the logic model regarding data gathering skills and diagnostic hypotheses. There were two themes, 'Strategies' and 'Being flexible and open-minded about diagnosis', identified that suggested students use different data gathering strategies or approaches to eCREST. The other themes suggested the variety of other factors that might influence students' approaches to gathering data, 'eCREST's influence on clinical reasoning skills', 'Knowledge gaps', 'Confidence and uncertainty' and 'Students' engagement with eCREST'. Based on this data I referred back to the students quantitative data collected during the Think Aloud task to explore how students actually approached the cases in eCREST e.g. what questions they asked and their diagnoses. This approach led me to develop four typologies of approaches to gathering data, which will be described in the following section. The themes will be described further in relation to these typologies in **Section 8.3**.

Table 8-1 Overview of themes

Theme	Sub-themes					
Strategies	System-based questioning	Reflecting on questions they would ask in clinical practice	Being thorough	Being focused		
Being flexible and open-minded about diagnosis	Broad initial hypotheses	Staying open-minded	Fixating on primary diagnosis	Balancing likelihood and serious diagnoses	Keeping an open-mind in reality	
eCREST's influence on clinical reasoning skills	Checklist	Relevance of questions	Guided information gathering	Thinking of alternative diagnosis	Prompted to start think about diagnoses earlier	Time to reflect
Knowledge gaps	Generating hypotheses	Misinterpretation of information	Examinations and managing the patient			
Confidence and uncertainty	Lack of confidence	Clinical content of the patient cases	General Practice setting	Taking responsibility in a simulated setting		
Students' engagement with eCREST	Realistic	Attitude towards learning using eCREST	Interactivity	Format	Feedback	

8.3 How do students use their clinical reasoning skills in eCREST?

8.3.1 Identification of different data gathering styles from quantitative data

To explore my first aim of identifying how students use their clinical reasoning skills in eCREST, I investigated how these different skills were externalised via students' clinical reasoning behaviours (data gathering). Based on my learning from the feasibility trial, I focused on exploring differences in students' data gathering strategies as measured by key features one and two. I chose to focus on these two quantitative outcomes to create a typology, as these outcomes were the only quantitative outcomes from the trial to be significantly different (or approach significance) between groups, which suggested these may be particular behaviours that eCREST influences. The quantitative data indicated that students' data gathering strategies fell across two continua: relevance of questions asked (focus) and comprehensiveness of information gathered (breadth). Therefore, there were four possible different styles that students could demonstrate, which I describe in **Table 8-2**Error! Reference source not found..

One strategy I identified was being 'Focused'. This involved identifying most of the essential information available and asking few irrelevant questions. Others found it difficult to identify all of the essential information without also asking irrelevant questions, and took a more 'Thorough' approach to gathering

data. In clinical practice, this may not be feasible or acceptable to patients but this approach to gathering data may be expected of novices, given their limited experience and knowledge. Other students were relatively 'Succinct' in their approach. They asked for generally relevant information from the patient but did not gather enough information to identify all the essential information they needed to make an informed diagnosis. Some appeared to take a 'Broad' approach, in which they gathered a lot of information but still missed some of the essential information.

Table 8-2 Different styles of gathering data

Name of style	Description of styles for data gathering
Focused	Student identified essential information and asked for mostly relevant information. This strategy is efficient, as the student gathers all the key information without wasting time on unnecessary questions or tests.
Thorough	Student identified essential information but asked for a lot of irrelevant information too. Student is unlikely to miss key information but may waste time on gathering unnecessary information.
Succinct	Student did not gather enough information to identify all the essential information they needed to make an informed diagnosis but most of what they asked was relevant. This strategy saves time on not asking irrelevant questions but it risks missing key information.
Broad	Student gathered a lot of information but did not identify enough essential information to make an informed diagnosis. This strategy would be more common if the student had little knowledge of the case and in simulated settings where an unlimited list of questions is available for students (such as in eCREST).

I labelled the different styles of data gathering of students in the Think aloud study based on their quantitative data gathering characteristics reflected in key features one and two (Error! Reference source not found.**3**). I then looked at their performance on these key features by patient case, as the results from the validation of the key features suggested that there might be differing levels of difficulty or case structure between the patient cases that may have influenced scores on the key features. Indeed, I found that performance on the key features varied by case. For example, in case one the highest percentage of relevant information asked was 79%, whereas for case two everyone scored above 85%.

Table 8-3 Think Aloud students' demonstrated strategies according to quantitative data collected by eCREST

Patient case	Participant number	Strategy	Relevant information gathered (key feature 1)	Essential information identified (Key feature 2)
1	6	Focused	79.20%	70.00%
1	15	Focused	75.90%	85.00%
3	3	Focused	80.70%	75.90%
3	7	Focused	92.86%	72.41%
4	10	Focused	82.80%	65.50%
4	14	Focused	86.70%	75.90%
2	1	Focused	87%	76.90%
1	11	Thorough	71.10%	100%
4	4	Thorough	68.90%	79.30%
2	9	Succinct	90.90%	69.20%
1	5	Succinct	75.00%	55.00%
2	2	Succinct	85%	69.20%
3	12	Succinct	75.90%	65.50%
4	8	Succinct	87.50%	51.72%
2	13	Broad	80.00%	69.20%
3	16	Broad	71.90%	65.50%

Note. Red indicates that a student's score was in the lowest quartile for that patient case. Green denotes that a student's score was in the top three quartiles for that patient case.

In Error! Reference source not found. I classified the students who scored in the lowest quartile for each case to have scored worse on that key feature. I colour coded the table red for those who scored in the lowest quartile for each case and colour coded the table green those who scored in the top three quartiles on the key features for that case. Those whose scores were in the top three quartiles for key feature one and two, were closest to demonstrating the ideal strategy of being 'Focused' ($n = 7$). Those whose scores were in the top three quartiles for key feature one but in the lowest quartile for key feature two, compared to others in that case, were closest to demonstrating the 'Succinct' strategy ($n = 2$). Those whose scores were in the lowest quartile for key feature one but the highest quartile for key feature two, were closest to demonstrating the 'Thorough' strategy ($n = 5$). Those whose scores were in the lowest quartile for key feature one and key feature two, were closest to demonstrating the 'Broad' strategy ($n = 2$). Strategies appeared to be reasonable evenly distributed across patient cases. While I expected that different patient cases and symptoms might influence data gathering style, this suggests that other factors, such as the design of eCREST and other factors, might also influence how a student approaches gathering data.

The following sections compare the themes identified from the qualitative data, collected in the Think Aloud and interviews, to describe and define each data gathering strategy.

8.3.2 Theme 1: Strategies

8.3.2.1 System-based questioning

Students mentioned using specific strategies, which they had learned in medical school, to help them structure how they gathered information from the patient. A system-based approach was one of the most common ways of approaching the cases. This approach involved ruling out the causes of the symptoms one system at a time e.g. cardiac then respiratory causes. I expected use of this strategy given that SA's curriculum takes a system-based approach. Students displayed evidence of using system-based strategies when using all data gathering styles apart from those who took a 'Broad' approach. It was most evident in those who were 'Thorough', perhaps suggesting that this is a good strategy for detecting essential information to make an informed diagnosis but also may lead to some unnecessary questions being asked.

"I really should rule out a cardiac thing actually, so I'll do that now.
Um I think I normally like to sort of focus on a system, so do almost a
respiratory um thing and then move on to cardiac." [Participant 4 –
Thorough]

8.3.2.2 Reflecting on questions they would ask in clinical practice

Many students also considered how they usually gathered information from a patient in clinical practice to guide their history taking. This showed that students were applying skills both learned from formal teaching and from clinical practice, even though their clinical experience was relatively limited. It also suggested that the clinical reasoning skills they demonstrated in eCREST were a good proxy for their approach with a real patient, as they acted as if they were in a real-life consultation scenario. Students who took a 'Focused' approach and

269

a 'Broad' approach reported during the Think Aloud using this technique for selecting questions more than the other groups. However, those who used a 'Focused' approach demonstrated evidence that they thought about what they wanted to ask before seeking a question in eCREST, so demonstrated the rationale for their questioning. Whereas, those who took a 'Broad' approach showed that they were more directed by eCREST and did not necessarily have a rationale for their questions; they scanned the questions available in eCREST before thinking about whether they would ask that question in clinical practice.

"Have you had an injury to your chest? Do I ask it, do I not? I've never asked it in real life to a patient. Mm, mm. No. I'm gonna leave it. That doesn't make you cough up sputum. If you had a really bad injury to your chest, you'd cough up blood and would it be a short onset cough.." [Participant 16 - Broad]

"So, thinking about ... I would always ask, 'What do you think is going on?' to the patient – but I'd ask that after I've had my ... so it doesn't bias what I think. Um, so, if I think this is cardiac related – which I do – I would ... my thought process is then to move into risk factors for cardiovascular disease." [Participant 1 – Focused]

8.3.2.3 Being thorough

Students whose strategy was 'Thorough' and 'Broad' reported in the Think Aloud that they wanted to be thorough and ask all of the questions available, seemingly regardless of relevance to the patient case. Both groups showed during the Think Aloud and interviews that they were aware they were asking irrelevant questions but only those who were 'Thorough' showed that they recognised that their strategy was taking them a long time and was potentially inefficient. One of the reasons those who were 'Thorough' asked many questions was to reassure themselves that they had not missed something serious.

270

"So these questions I guess aren't like very helpful but because they're there I'm going to ask them anyway." [Participant 11-Thorough]

"RES :..Can I click through all of them [questions]? INT: If you want to. RES: If I want to [laughing]. Okay. INT: It's up to you. RES: Oh okay, no I mean because of all the… It would be useful." [Participant 13 – Broad]

By contrast, students whose approach was 'Focused' and 'Succinct' reported in the Think Aloud they wanted to limit the number of questions they asked and expressed they did not want ask too many questions. They seemed more aware of the importance of using time efficiently in real clinical practice.

"Erm… I think that's probably, I don't really want to ask any more of these I mean partly because I feel like I don't, I want to be, I want to be focussed, so I'm not really going to ask the rest of these which are potentially not that related.." [Participant 10 – Focused].

Those who used a 'Succinct' style reported either feeling that they had asked enough questions or recognised that they did not ask enough questions in the interview. Indeed, one student who was 'Succinct' in their approach reported that they had limited their questions too much and reflected that they should have used a learned strategy, such as taking a system-based history to help them structure their questioning.

"I found choosing the right questions in the initial history presenting complaint quite difficult. So, maybe I should have not tried to limit myself to a specific number and asked what I thought was actually appropriate. And in a system based, rather than just all over the place. So, system based questioning. In a history presenting complaint, rather than just clicking randomly." [Participant 2-Succinct]

271

8.3.2.4 Being focused

In the interviews, just under half of the students in the 'Focused' group felt that their approach to history taking was quite "focused", they felt that they had gathered the right amount of information and seemed aware that they had a good strategy. The other half expressed that they felt they needed to be more "focused" and "structured". Thus, students' insight into their data gathering strategies varied and was not always reflective of how they actually performed according to the key features I identified.

> *"So ask questions that are focused if you have a differential in mind. Because I did this when I was at GP the other day, and I didn't actually, I had an idea of what it could be, but I didn't actually ask enough questions about that thing. So if I thought it's interstitial lung disease I should have asked about exposure." [Participant 14 - Focused]*

In the interviews, those who were 'Succinct' in their data gathering style mostly reported they needed more "focus". This was surprising since they asked mainly relevant questions and relatively few questions overall but they may have been referring to the fact that they did not identify all the essential information they needed to reach an informed diagnosis. Those who were 'Thorough' and 'Broad' in their approach to data gathering showed mixed awareness that they needed more focus when gathering data. This further demonstrated that students have differing levels of insight into their performance.

> *"I think I asked questions reasonably logically or like with a logical flow. But I think I could be more concise. Cos, I just kind of ask everything just in case." [Participant 16 – Broad]*

272

"Tended to ask too many unnecessary investigations…Picking up the red flags. Being more succinct in questions and doing it faster."
[Participant 11 – Thorough]

8.3.3 Theme 2: Being flexible and open-minded about diagnosis

8.3.3.1 Broad initial hypotheses

Students of all data gathering styles demonstrated evidence that they attempted to keep an open-mind about diagnoses from the beginning of the task. When generating their initial diagnostic hypotheses, most students recognised it was important to place the most likely and most serious diagnoses in their differential diagnosis.

"So most likely or most yeah, so the way I think about the question is, because there's like most likely and there's like what you're most concerned about. Like I don't know how common lung cancer is, but that's probably what I'm most concerned about. So I'm going to put that as my top one, but I don't know if it's the most likely." [Participant 14 - Focused]

Those whose data gathering style was more 'Broad' were poorer at considering the most serious diagnoses compared to the other groups. In the Think Aloud and interviews, students who took a 'Broad' approach showed evidence that they might not have included serious diagnoses like heart failure because they had so many diagnoses in mind initially that they found it difficult to focus.

"And in terms of preference, what's more prevalent, lung cancer or TB? Yeah, let's leave it like that. But if there was a six and seven, I'd want to put ACS [acute coronary syndrome] because I'd want to rule out in real life." [Participant 16 – Broad]

273

Those whose approach was 'Focused' had varied thoughts during their reflection and in the post Think Aloud interview on whether they were open-minded about the diagnoses from the beginning of the task. One student who took a 'Focused' approach felt that they began the task considering a range of diagnoses, whereas, another student who took that approach felt that if they had had a broader range of diagnoses initially it would have helped them to stay more open-minded about diagnosis.

"I think at the beginning I had a broad range of differentials, as in before I started taking the history, history and I think I probably thought about, mm, maybe not thought about most of them but I think I had quite a few of them I guess." [Participant 10 - Focused]

"I should have had a broader initial differential diagnosis list for, er, tired all the time. Um, maybe I should have started with background information? And I think that might be because I'm currently placed in A&E, where no patient comes to you without some kind of background – they're always triaged by someone else, who's non- ... who's not a doctor. Um, in GP, I think that's very different. So, that's ... I think that's a slight bias, on this occasion." [Participant 1 – Focused]

Most students who took a 'Thorough' and 'Succinct' approach felt they should have had a broader initial diagnosis.

"So, it's … it's difficult to rule out a diagnosis completely. … um, but to have a … a blank … to be um open minded and think from the very beginning what might be wrong with the patient." [Participant 12 – Succinct]

8.3.3.2 Staying open-minded

Keeping an open mind about diagnoses during the task seemed more challenging than at the start of the task. Those whose data gathering style was 'Succinct' and 'Thorough' seemed to be comfortable selecting the most likely

diagnosis, but selecting their third, fourth and fifth diagnoses (the less likely diagnoses) seemed more challenging. They struggled to widen their diagnosis and think of a range of likely diagnoses and were more reassured by the absence of symptoms than those who were 'Focused'.

> *"Okay erm, I will put in … I'm just going to put in, arbitrarily, asthma, err … ischemic heart disease, err … I don't think she had left heart … okay." [Participant 8 – Succinct]*

> *"Can you describe your cough? Okay so two three weeks that makes the kind of short term ones more likely as opposed to viral. Makes the post nasal drip less likely. Makes the pneumonia a bit more likely. He's not unwell though so actually pneumonia isn't terribly likely. It makes me less worried that it's kind of TB, makes me less worried that it's cancer." [Participant 11 – Thorough]*

Those who used a 'Focused' style displayed evidence that they considered serious diagnoses but reported feeling uncomfortable about selecting unlikely diagnoses. Indeed, one student felt they were being "dramatic" by including unlikely but serious diagnoses. Those who had a 'Broad' approach showed little evidence that were keeping an open-mind throughout the consultation about diagnoses but in their reflection they also expressed feeling uncomfortable investigating serious but rare conditions.

> *"Lung cancer was at the top of my list as it would be the most worrying diagnoses. However, after examining her the findings of mild pyrexia and right mid zone crepitation's pointed to a lower respiratory tract infection. It would be important to still follow her up for lung cancer in case the chest infection was super imposed." [Participant 3 - Focused]*

8.3.3.3 Fixating on primary diagnosis

Cognitive biases influenced all styles of data gathering, as some students across all styles became fixated on a diagnosis and would

consequently seek information to confirm that diagnosis rather than seek information to refute it and did not consider what other diagnoses would also be relevant given the symptoms. However, only one student who took a 'Thorough' approach showed awareness of this during the task and made a conscious effort to seek other information, but consequently may have asked too many irrelevant questions.

> *"I really should rule out a cardiac thing actually, so I'll do that now. Um I think I normally like to sort of focus on a system, so do almost a respiratory um thing and then move on to cardiac. Um although I shouldn't get too into confirming about, I'll just ask about any other symptoms." [Participant 4- Thorough]*

Those who were 'Focused', 'Thorough' and 'Succinct' all reflected in the interviews that they fixated on one or two diagnoses and they tried to make all the information they gathered relate to those diagnoses, while ignoring other relevant diagnoses. Those who used a 'Broad' strategy did not report fixating on a diagnosis but this could have been because they were unsure of the most likely diagnosis.

> *"Well I missed the PE bit. So risk factors for PE. And considering everything the patient has said and I think just not trying to make diagnosis fit like the COPD that I was trying to make her fit." [Participant 7 - Focused]*

> *"But often I ask confirmatory questions and be like, and just exclude things that I just know weren't on my differential and so my differential didn't really change." [Participant 11 – Thorough]*

Some in the 'Succinct' group even acknowledged that once they had decided on a diagnosis they stopped investigating other causes. However, they showed awareness that in reality patients often do not present with many red flag symptoms and have a typical presentation of a disease; thus, they

understood why it was important to keep an open-mind in this clinical context. In contrast, one student who had a 'Succinct' data gathering style felt that they were "flexible" with their diagnosis and changed their differential diagnosis many times when prompted to do so by eCREST. They felt this helped them keep an open-mind and carry out multiple lines of questioning.

> *"I think the fact that I kind of was able to change my differential diagnosis every time I was given the option to, and that wasn't just because of giving the options, like giving the options to you but it was because I wasn't trying just like take one line of questioning and run with it, because, erm, he could have been breathless and I could have made it all respiratory based and I kind of wanted to find out a bit more about. I think being flexible with my diagnoses, erm, that when I first came in I was convinced it was diabetes and obviously it wasn't. I think being flexible with diagnoses." [Participant 9 - Succinct]*

8.3.3.4 Balancing likelihood and seriousness of diagnoses

Across all the data gathering styles, students in the interviews reported that they needed to be better at considering and investigating the "likelihood and emergencyness" of diagnoses. They felt they needed to ensure that they ruled out the most dangerous diagnoses first, even if certain aspects of their history made them unlikely. Students were particularly concerned that they had not considered and investigated lung cancer. Those who were 'Broad' in their approach least commonly expressed this view.

> *"I guess also just remembering that common things are common and things like postnasal drip being higher up on the differential list than something like interstitial lung disease … So I'll put putting more common differentials first … Erm, oh, yes and they asked as well, I guess is, erm, putting lung cancer like quite high up and so you could rule that out, erm, if that's the most urgent one to rule out … And something to improve I guess, erm, asking all questions about, erm, lung cancer before like getting distracted and moving onto others."*
> *[Participant 6 - Focused]*

"what do you need to improve. Keep an open mind. And also, perform tests to rule out the less likely diff… less likely diagnosis, differential diagnosis. As well as confirming the most likely."
[Participant 2 – Succinct]

"..I should definitely have been putting it [cancer] higher. So um weigh up likelihood, also I'm, this is my assumption, I'm probably right, it probably is more likely than the other things anyway statistically. The likelihood of a differential with the importance of ruling it out." [Participant 4 – Thorough]

8.3.3.5 Keeping an open-mind in reality

While students reported that they were aware of the need to investigate unlikely diagnoses, two students, one in the 'Focused' and one in the 'Broad' group felt that this might be unrealistic in clinical practice. They felt that in reality they might not consider serious conditions if a patient only presented with a few symptoms because it would be so unlikely that they actually had a serious diagnosis. Indeed, one student who had a 'Focused' style reported in the interview that their consultant would not "entertain" the consideration of a rare diagnosis. However, they acknowledged that this was perhaps more of an issue in secondary care than primary care. The student in the 'Broad' group had concerns about the efficiency of approaching a consultation if they were to keep an open-mind throughout.

"Um, I, I appreciate that ... what they've said about, um, all of their diagnosis, but this ... In real life, what happens is: you make a working diagnosis, and everything else is left behind – you don't continue those, generally. There might be some things you safety net, but by and large, when it's clear cut – as that, as that was – you would almost take that, go with it, do a few things, just to be sure."
[Participant 1 – Focused]

"Difficult to force yourself to include things that you aren't really entertaining the idea of seriously. Um, and you know, certainly ... Again, I'm biased, but where I am – and I know this is a GP situation, but in A&E, if you said, 'Oh, my fifth diagnosis of this patient is' – I

*don't know, um – 'Addison's Disease,' I think you'd, you know, be
laughed out of the department, realistically."* [Participant 1 – Focused]

*"But I guess it's just getting a balance between kind of having those
differentials and shaving them down in your head. And having like a
few differentials and ignoring anything else the patient says and just
gunning for those differentials."* [Participant 16 – Broad]

8.3.3.6 *Quantitative data relating to flexibility and open-mindedness*

Given that flexibility in diagnosis and keeping an open-mind about
diagnosis came out as an important theme in the qualitative data from the Think
Aloud and interview study, I investigated these themes in their quantitative data
collected during the Think Aloud task. I measured flexibility in diagnosis by
calculating how many times a student changed their diagnosis during their
patient case. I measured open-mindedness about diagnosis by calculating
whether they included the most serious diagnosis in their initial and final
differential diagnosis. I also measured the percentage of appropriate diagnoses
selected in their initial and final differential diagnosis. **Table 8-4** presents the
quantitative data outcomes relating to flexibility and keeping an open-mind.

For the number of times they changed their diagnosis and the
percentages of appropriate diagnoses, I assigned students as scoring worse if
they scored less than or equal to the lowest quartile value for that patient case.
For the outcome of selecting the most serious diagnosis, I scored students as
performing poorly if they did not select the most serious diagnosis. However,
due to differences across the patient cases in difficulty and symptom
presentation, I found that scores were relatively similar within patient cases for
some of the percentages and selection of serious diagnoses. For example, for

patient case one everyone identified lung cancer as an initial diagnosis and in patient case three all students who did that case identified 66.7% of appropriate diagnoses. Therefore, if three or four students scored the same percentage or all answered 'yes' or 'no' for one patient case I designated their performance as 'standard', as their answers were likely influenced by the difficulty of the case. I colour-coded **Table 8-4** to indicate performance. In the table, red indicates that for that patient case, the student scored worse than others did. Orange indicates that performance was standard for that case. Green denotes that the student performed better than others (in the top three quartiles) for that patient case.

Table 8-4 Indicators of flexibility and keeping an open mind about diagnosis for each Think Aloud student by data gathering style

Strategy	Patient case	Participant number	Number of times changed diagnosis	Selected most serious diagnosis in initial	Selected most serious diagnosis in final	Percentage of appropriate diagnoses in initial differential	Percentage of appropriate diagnoses in final differential
	4	10	4	Yes	Yes	66.70%	83.30%
	1	6	2	Yes	Yes	33.00%	66.70%
	1	15	3	Yes	Yes	50.00%	66.70%
Focused	2	1	3	No	Yes	42.86%	71.40%
	3	3	3	No	Yes	50.00%	66.70%
	3	7	4	Yes	No	66.70%	66.70%
	4	14	5	Yes	Yes	83.30%	100%
	1	11	3	Yes	No	50%	66.70%
Thorough	4	4	2	Yes	Yes	83.30%	83.30%
	1	5	3	Yes	Yes	33.00%	83.30%
	2	2	1	No	Yes	28.60%	42.90%
Succinct	2	9	3	No	Yes	57.10%	57.10%
	4	8	2	Yes	Yes	83.30%	83.30%
	3	12	5	Yes	No	33.00%	66.70%
	2	13	4	No	Yes	42.90%	42.90%
Broad	3	16	3	No	Yes	83.30%	66.70%

Note. Red indicates that a student's score was in the lowest quartile for that patient case. Orange indicates that student performance was standard for that case. Green denotes that a student's score was in the top quartiles for that patient case.

Those who used a 'Broad' approach may have had poor flexibility, as students using this approach were among the most likely to keep their diagnosis the same. They were also not very open-minded about the diagnosis, as some students failed to identify serious diagnoses initially and performed poorly on identifying appropriate diagnoses in their final differential diagnosis. The use of a 'Thorough' approach may also have indicated poor flexibility, as students using this approach were also among the most likely to keep their diagnosis the same. This style was also not particularly open-minded, as a student removed the serious diagnosis from their final differential diagnosis. The use of a 'Succinct' approach seemed to indicate a more flexible approach to diagnosis than the 'Thorough' and 'Broad' styles, as students using this approach were more likely to change their diagnosis. They were open-minded, as they were relatively good at selecting the most serious diagnoses but they identified few other relevant diagnoses in their differential diagnosis compared to other styles. Those who were 'Focused' showed they were the most flexible in their diagnosis, as they changed it the most times. They were less open-minded in the beginning of the task but tended to include the most serious diagnoses, as well as other relevant diagnoses in their final differential diagnosis.

8.4 How do different factors, such as the way eCREST is designed, students' confidence, and knowledge gaps, influence clinical reasoning skills?

8.4.1 Theme 3: eCREST's influence on clinical reasoning skills

8.4.1.1 Checklist

Some students in the interview reflected that the question lists had reminded them to ask a question that they might have otherwise forgotten to ask. This was most evident in students who used a 'Focused' approach (and by a student whose data gathering style was 'Succinct'). During the Think Aloud, those that were 'Focused' displayed evidence that they thought of their own questions first then used eCREST like a checklist to confirm they had not missed anything.

> *"So there's nothing else I really want to ask, except maybe was your home heated by an open fire or a wood burning heater, because I guess she's from Ethiopia so as a child, I know that's a major cause. If this doesn't prompt it I don't think I would have actually remembered, would have actually asked this to a patient but sure."*
> *[Participant 14 - Focused]*

During the Think Aloud, all students showed evidence that they used the diagnosis list in eCREST to help them check if they missed any serious diagnoses and it became a prompt for them to check if they had gathered all relevant information for their diagnoses.

> *"Okay so there's a lot of things [diagnoses] here. I'm just going to… I think I'll go through all of them. So I've ranked them from the most likely, so Addison's disease, that's probably quite rare actually I think…" [Participant 9- Succinct]*

8.4.1.2 Relevance of questions

Some students in the interview reflected that eCREST helped them to understand why they were asking questions and to ask questions that were relevant to their hypotheses. This was particularly expressed by those students who had a 'Focused' data gathering style who identified that some questions given in eCREST were irrelevant to the case.

> *"it did make me feel that I shouldn't keep going until I do every single question, um, because it was quite clear that there were a few which wouldn't have helped." [Participant 1 - Focused]*

> *"yeah I thought it was good, like just the way that you have to ask them, it lets you ask a certain number of questions and then you always, it makes you like re-evaluate your ranking of diagnosis because then you actually have think about the questions and why you're asking them in the first place, so yeah." [Participant 10 – Focused]*

In the interview, those who had a 'Succinct' approach reported mixed views on whether eCREST steered them towards pertinent questions. Some felt that eCREST helped them to identify questions that were specific to their hypotheses and stay focused. However, another student felt the list of questions in eCREST "biased [their] questions to be random".

> *"it was good that it gave you all the different drills so you could pick from them and it reminded you about ones that I wouldn't have remembered before." [Participant 5- Succinct]*

Students who approached the cases by being 'Thorough' also had mixed views on whether eCREST helped them to ask pertinent questions. Some felt they asked random questions, whereas some felt their questions would have been even more random had it not been for eCREST.

"Um, but it … the problem was I didn't really think about what questions I was going to ask. Instead I looked at what questions were there and I chose whether or not to ask them." [Participant 2- Thorough]

"I wonder if I would have asked some more kind of random questions, because some odd differential would have come to my head. And I think also going oh just judge these ones as my sort of prompts." [Participant 4- Thorough]

Those who had a 'Broad' data gathering style commented in the interviews that there was "pressure to ask only pertinent questions" in eCREST. This indicates that they may have been aware that eCREST was directing them to ask only relevant questions but they found this uncomfortable to do in practice, perhaps because they feared missing key information or because they felt they were being tested.

"Um, yeah there were um, I felt much more under pressure to you know ask like sensible questions and not just ask everything. [Participant 16 – Broad]

8.4.1.3 Guided information gathering

Some students in the interviews reflected that eCREST helped them to organise the way they gathered information, as it only allowed students to ask six questions at a time, and required students to reflect on their diagnosis after every six questions. Those who used a 'Succinct' data gathering style referred to the way that eCREST helped them to be more succinct and "streamline" their thoughts by asking students to think of questions in sets of six. It seemed that by forcing students to "chunk" their history taking into sections, eCREST helped them to think about structuring their history taking in sections by red flag symptom or system.

"I guess it's good because it makes you streamline your thoughts regarding diagnosis after you have limited information available and I think that probably helps time management within GP settings, because it's making you streamline your questions, like 6 questions, 8 questions, 8 investigations, erm, and I think as a GP that would help with time management and just getting to the point of having to make sure that you get the diagnosis that you need and the time available." [Participant 9 – Succinct]

Students who were 'Broad' found it useful to visualise their diagnostic hypotheses regularly, which helped them to "hone" their questions towards their diagnoses.

"I think it was useful even for me to just like, to see when I'm taking the history, I think that just that, I don't really write down all the five, the top five differential diagnosis when I'm taking history but more like this is what I'm worried about, and then I want to ask other questions that are related to this, this, and this, and I want to rule that out first, and then I'll look at other things, before that, which I think it really helps me to, erm, to put things, yeah. Yeah, just better at visualising it, and organising it." [Participant 13 – Broad].

Some felt during the task that eCREST prompted them to ask questions that they would not have normally asked and saw this as having a negative impact on their data gathering. Students who had a 'Focused' and 'Succinct' style reported during the task and in the interview that they would usually ask open questions, which they felt would give them more information from the patient than the closed options available in eCREST.

"It's just a bit weird because like because you've got those specific questions to ask like, and it's not. It's quite useful to make you think about the questions but because I don't ask questions in that, I feel like my own style is quite different to the way it's set out here. Like I'd be quite, I'd be quite like open with the patient. I'd be like tell me more. And then I'd be able, I'd have some better idea, I'd have a better timeline of the things. But it's yeah. We'll see." [Participant 15 – Focused]

"Um, and I guess maybe there were some questions that I thought were a bit limit … like a little bit closed or a bit …like limited."
[Participant 12- Succinct]

During the task, one student who had a 'Focused' style seemed to recognise that eCREST was starting to influence the way they would usually gather data, so they purposefully tried to answer the questions before looking at the list of options available in eCREST.

"Well, it's probably useful to try and think about this before I look at the list." [Participant 1 – Focused]

Some students whose approach was 'Succinct' also recognised that eCREST diverted them from their usual strategy, as they reported during the Think Aloud that they had should have "done questions they wanted to do first" and felt eCREST forced them to make a diagnosis.

"… when actually, if I'd just stuck to my structure … it would have made a lot more sense. Like normally, you know, normally when you need to know about someone's cough you ask the basic questions then you move on to other things, and … I think I just panicked and was like better cover cardio and respiratory in the same questions. So, that did irritate me a little bit." [Participant 8 - Succinct]

Students who were 'Broad' and 'Thorough' expressed that they felt the constant reviewing of the patient meant that they would sometimes forget what they had asked and they thought in too much of a "modular" way; they felt they were more led by the questions in eCREST than the patient's response and at times responded negatively to the structure they were given by eCREST to gather data and test their hypotheses.

"Um, but I think it can … it did perhaps make me think in that kind of modular way...o, each time I only considered the six questions that

had been before. And forget about what had happened before that.
So, like less of a kind of continuous set of questions and more like,
oh, in the last six questions she said she like didn't have a fever…
and she had this. So, and I just used that information and I just need
to think back uh…to the questions that I've been … asked before."
[Participant 16 – Broad]

"Um and I think in a consultation I tend to try and slightly go along
with the patient and take the opportunity to ask questions. So
sometimes say I'd ask about pets, if they brought up pets even
though in my system, in my head pets would come in social history.
Um so I suppose I don't know, maybe I was slightly more, I feel like
maybe that's unfair but slightly more formulaic than I might have done
it." [Participant 4 – Thorough]

During the task, students who were 'Focused' in their strategy also expressed that they wanted to ask questions that were not available to them. This could suggest that students who were 'Focused' might have had more ideas of questions that they wanted to ask than those who used other styles. However, as one student pointed out in the interviews, it could suggest that the questions they would usually ask were less relevant than ones available in eCREST.

"So she's coughing a lot at night. I'd like to ask her also how bad the
pain is. Like she didn't really expand on that very much. Is that a
question I can ask? Not really. Okay." [Participant 3 – Focused]

"And then yeah I guess just having the list of pre-selected questions
meant that some of the like standard questions that I always ask just
weren't there. But that probably just goes to show that they're not as
important to ask as I think they are." [Participant 3 – Focused]

8.4.1.4 Thinking of alternative diagnosis

In the interviews, some students who had a 'Broad' approach commented on how being prompted to review their diagnosis had allowed them to keep an open-mind because it made them think about more than one or two diagnoses. Students across all different data gathering styles, except those who

were 'Thorough', felt that eCREST helped them to consider alternative, less likely diagnoses and to justify their diagnoses. However, some who had a 'Succinct' data gathering style also felt that they arbitrarily selected diagnoses that were on the list, as they found it difficult to think of other likely diagnoses.

"the fact that it makes you reconsider [typing], erm, it makes you reconsider your diagnosis, your diagnosis after asking questions, asking a set number of questions is good practice for reality [typing], when you should be doing that but you probably don't." [Participant 10 - Focused]

"Erm … I … thought that six choices were quite a lot to rank for differential diagnoses. I think in practice they always ask you to write as many diagnoses as possible, erm … but I think I'd … at some point I was kind of making up the things that I think I would have forced in." [Participant 8 – Succinct]

8.4.1.5 Prompted to start thinking about diagnoses earlier

Students who had a 'Focused', 'Succinct' or 'Broad' approach reflected in the interviews that reviewing their diagnoses early in the consultation made them think of diagnoses at an earlier stage than they usually would. Some students reflected that they usually only think about diagnoses at the end of a consultation and being prompted to think about them continually in the consultation helped them to "streamline" their questioning towards those diagnoses.

"I guess that you're kind of, you're expected to commit to things quite early on and then adjust them. That's quite good instead of just sort of going in to questioning thinking that anything could be wrong. You already start to think about narrowing things down. So that's quite good." [Participant 3 – Focused]

"Um, so, yes, because usually I don't really think about differentials so early on in uh consultation um … so this encouraged me to um ruling out different diagnoses um from a very early point. A very early response to this might be useful I guess." [Participant 12 –Succinct]

Those who used a 'Broad' approach also felt uncomfortable with thinking about their diagnoses so early in the process because they felt it forced them to narrow down their diagnoses too soon and led them to not investigate important information. However, they acknowledged that the structure given by eCREST to think about diagnoses earlier was perhaps more efficient than their usual strategy, as it prevented them from asking all the questions.

> *"Um, made … made me think of my differentials much earlier in the whole process which made it more confuse … oh not confusing, more challenging, because you feel the need to ask only pertinent questions. Which in one sense can be dangerous because then you think, if a question is not gonna be pertinent to those differentials you don't ask it. Whereas before I would have just asked everything anyway and then I wouldn't have missed anything. But also, I would have taken triple the time. Um, yeah, so that … that … but I guess that's clinical judgement isn't it." [Participant 16 - Broad]*

8.4.1.6 Time to reflect

Students who took a 'Focused' and 'Succinct' approach also reported that they found it useful that eCREST gave them time to pause, think and reflect on their questions and diagnoses. They reflected that in real life it would not be possible to have an "awkward gap" to think about your differentials.

> *"So it's nice to just click the questions, and then spend five minutes thinking about it. Um and I think when you're actually seeing patients there's emphasis on it being slick." [Participant 14 – Focused]*

> *"..in your mind its quite a lot, but it's nice to have the time to just think about it." [Participant 12 – Succinct]*

290

8.4.2 Theme 4: Knowledge gaps

8.4.2.1 Generating hypotheses

Students who had a 'Succinct' data gathering style showed that when they were starting the task and thinking about their initial diagnoses they had gaps in their knowledge. Some students did not know about some of the diagnoses presented on the list of diagnoses or had little understanding of what the risk factors and symptoms were relevant for the different diagnoses. However, this did not appear to stop some students selecting that diagnosis.

"Erm, next I'll probably go for, I'm not sure what pertussis is, I can't remember." [Participant 5 - Succinct]

"Hypothyroidism could be a big one, makes you very tired, erm, as well but I don't know any other symptoms." [Participant 9 - Succinct]

8.4.2.2 Misinterpretation of information

During the Think Aloud, there was evidence that students across all data gathering styles had misinterpreted the patient's symptoms (the clinician who checked over the transcripts identified these mistakes). This was most evident in students who had a 'Succinct' and 'Broad' style. Those who used a 'Broad' approach verbalised some incorrect interpretations of the symptoms presented. Those who used a 'Succinct' data gathering style also failed to recognise that finding no symptoms for a suspected diagnosis does not necessarily rule out a diagnosis. For example, one participant showed awareness that the presence of haemoptysis would make TB more likely but failed to recognise that the absence of haemoptysis would not exclude TB. Some students who had a 'Thorough' style struggled to pick up on key red flag symptoms for lung cancer,

such as appetite loss and duration of cough. Students who were 'Focused' verbalised that that they felt they needed to "revise" to improve their knowledge.

> *"Uh I'm really unsure now. Um I feel like it's less and less likely it's lung cancer. I'm going to put ILD second. Why? She had bilateral inspiratory crackles at the bases. Makes me think that ILD is more likely, or maybe you should go away and revise some respiratory medicine." [Participant 7 – Focused]*

8.4.2.3 Examinations and managing the patient

During the Think Aloud, the GP registrar (JS) noted that students demonstrated knowledge gaps when they ordered bedside tests and did physical examinations. Students across all data gathering styles ordered irrelevant tests. None of the students were aware of the purpose of the Patient Health Questionnaire (PHQ-9) and yet those who were 'Thorough' and 'Succinct' commonly selected this test, despite not being able to interpret it. Many students struggled with managing the patient because they did not know the appropriate referral guidelines. Students who were 'Broad' and 'Succinct' seemed to struggle to manage patients more than those who used other styles. They also displayed evidence during the Think Aloud that they lacked knowledge of how much primary care is involved in the diagnosis and management of patients when secondary care is involved.

> *"How would you like to refer. I mean it's unlikely to be cancer, no, no, multiple myeloma… Multiple myeloma is an urgent… I would actually refer him to… Oh, wait, no… interesting. It is not in GP practice, non-urgent referral, I mean this should be… I mean given his symptoms, erm, I'm sure there was a question about back pain actually because one, because of what I'm thinking about, I mean all these things are probably not that urgent, except for multiple myeloma, multiple myeloma is only urgent referral that I'll need to do." [Participant – 13 - Broad]*

"I don't think she's a two week wait. She hasn't got … she … I can't remember what the guidelines were; she has a chronic cough, she's had it for … time. Gosh, unless you refer everyone who … I think you do, I can't remember. She's got a really bad cough, and she's a smoker, she's had it for two years, and I feel like that's like a red flag, like you need to refer someone to … erm … secondary care. But all her tests suggest that she's got … COPD. Okay. Goodness me. I don't think she's a two week wait; I feel like she's a two week wait, I don't know. I can't remember. I feel stupid. Erm … okay, let's not urgent refer her." [Participant 8 -Succinct]

When students evaluated their learning from eCREST, they were able to identify several knowledge gaps. Students across all data gathering styles reflected that they needed to understand what the appropriate referral pathways and safety netting procedures are for patients in general practice.

"At the end when I was like you know, you probably should have done the chest x-ray, I wasn't, I personally wasn't convinced. But then like, then I mean I'm talking about, don't talk against the judgment of clinicians. So I guess it is a fair thing and it probably should have been, maybe my personal knowledge then on the like pathways, the referral pathways for lung cancer, referral pathways for an unexplained cough need to be reviewed. So I think it did, it was good in that it highlighted right you need to read up on this kind of thing." [Participant 15 - Focused]

"Um although definitely could have I think, you know, I don't necessarily know my management as clearly, so I should have been specific to be like I want an urgent x-ray, which is probably I think the thing I was most concerned about that I missed. Um in my head I wanted it to be basically, because you know, on GP you can have the x-ray done that day normally and get the results within a few days. But I should have said an urgent." [Participant 4 – Thorough]

8.4.3 Theme 5: Confidence and uncertainty

8.4.3.1 Lack of confidence

Students who were 'Thorough' and 'Focused' lacked confidence in selecting a diagnosis initially, as they did not feel they had enough information to make a decision and felt that eCREST forced them to make a judgement

before they were ready. Students who had a 'Succinct' approach also described not feeling confident because they were concerned they were missing key information. However, they seemed more comfortable with being forced to make this decision by eCREST. Those who used a 'Broad' approach did not comment on feeling unconfident in their clinical diagnoses because they lacked information, possibly because they asked many questions and felt they had a lot of information.

> *"I'm not very confident on the order I've chosen, um and I feel like if I was seeing this patient I would have these as differentials and then try and, almost try to exclude each of those with my questions specifically, and obviously with examination." [Participant 4 – Thorough]*

> *"I'm dreading the fact that she's probably got cancer and I've missed it." [Participant 4 – Thorough]*

> *"What is my most likely diagnosis? Oh God. I find that tricky just to rank them at this stage but I guess that's okay." [Participant 7 – Focused]*

In the post Think Aloud interview, some students who were 'Succinct' felt that eCREST created uncertainty because the lists of diagnoses and questions "put a seed of doubt" in their minds and encouraged them to "second guess" themselves. Students who were 'Thorough' and were 'Succinct' both reflected that they felt they needed more confidence. They expressed that there is a tension between making confident decisions about what they think the likely diagnosis is and fearing the potentially serious consequences for the patient if they misdiagnosed them.

> *"I think maybe sometimes I felt obliged to change my… list even though I hadn't really elicited that much new information…Um, but you know, I think maybe I should just have more confidence. That's maybe just a personal thing. Maybe I should just have more*

*confidence in saying that, okay this is what I think and it's still
consistent you know." [Participant 12 - Succinct]*

*"You can give every investigation and then be sure, but actually
realistically when you're trying to think well I can't get every single
blood test in the world, this is the initial management, I'm not going to
give them. And trying to think of that. And trying to do it and being,
at the moment I feel I have to be quite brave, because you think well
what if I do miss something that's terrible? But then I suppose it's
easy to think okay, well what are the absolute terrible things? Make
sure that I don't miss those, so for example an x-ray would cover a lot
of bases in a sense." [Participant 4 – Thorough]*

8.4.3.2 Clinical content of the patient cases

Students who were 'Succinct' in their approach, reported feeling confused and "frustrated" that the patients in eCREST presented with very few red flag symptoms. Students who used other data gathering styles also felt confused at times because the symptoms suggested that the patients could have more than one problem, and they seemed unsure of how to manage this uncertainty.

*"Okay, any additional questions, I've got no idea what's going on
now, so I'd be better. When are you worst, and we try and rule out
heart failure….Okay this is a little bit frustrating, when did your cough
start?" [Participant 5 - Succinct]*

*"Hypertension really is the only thing to make him cough and loose
an appetite or without losing sleep, having a fever, loss of weight,
don't know. Don't know, conundrum. Or it could just be two things
superimposed upon each other." [Participant 11- Thorough]*

*"This is just kind of, it's a bit confusing because she might have more
than one problem. Like I don't think, like I think there's probably
some underlying COPD, but I'm still worried about lung cancer
because she smokes a lot, and she's had this cough, I feel like she's
had it for ages." [Participant 14 – Focused]*

8.4.3.3 General Practice setting

Some students who were 'Focused' felt that eCREST had reminded them of the reality of uncertainty in general practice. They identified that they needed to learn to be more comfortable with and accept uncertainty and they recognised the importance of safety netting in general practice to manage this uncertainty.

> *"Learning to be comfortable, comfortable with a degree of uncertainty and is important to GP and several other specialities. I will need to learn from more experience how much safety netting and investigation for other possible differentials is appropriate."*
> *[Participant 3 - Focused]*

8.4.3.4 Taking responsibility in a simulated setting

Some students who were 'Broad' in their approach felt that in reality they would not be responsible for some of the decisions they made in eCREST, particularly around managing the patient. Therefore, they found it difficult to make decisions on their own in the simulation and wanted to speak to someone more senior before making a decision. Students who were 'Succinct' felt that there was pressure to make a decision in eCREST but they found this useful and engaging. However, they treated eCREST as a "game" rather than feeling immersed in the reality of the situation.

> *"..felt like you're kind of trying to make a decision which is as you've a question to ask. And then once you've decided that is it, was it a useful question to ask, then you've got to look at the information you're given in the video. So, it's at every, every sc… every click is almost a decision. Um, so I, I found that quite … I thought it was much more exciting and I was kind of, it felt like a, a game, not a game but it felt like um, I don't know, it just felt more challenging than previous things."* [Participant 2 - Succinct]

Students who were 'Focused' reflected that in future simulations they wanted to "act like only [they were the] point of care for [the] patient" to help them get used to making decisions in uncertain circumstances.

> *"I think every time, every time I do a case, you know, like case of the month or cases like this, and certainly when I see real patients, like I'm on GP at the moment. Either a, you know, case of the month exercise or in real consultations I'm going to try and think about it, to approach it as though I was the GP seeing the patient alone, and you're their only point of care. Sort of they're, yeah patient alone and therefore fully responsible for them. Which forces you to really think carefully about differentials and things not to miss I suppose."*
> *[Participant 4 - Focused]*

8.4.4 Theme 6: Students' engagement with eCREST

8.4.4.1 Realistic

All students expressed that the general practice scenario in eCREST, in which patients had very vague symptoms, was realistic and representative of a patient they might see in general practice. 'Focused' students noted that learning these skills on wards with real patients would be better but acknowledged that this is not always possible.

> *"Like patients do come in with weird vague symptoms like that. So like in, yeah, do you place those patients that just come in with oh my finger hurts or like oh yeah I've got this cough and that sort of thing. So yeah and sometimes like I could definitely see John Roberts just being a patient that you'd see in a GP like on a like." [Participant 11 – Thorough]*

> *Cos, you often get patients in GP who end up having a chest infection but aren't bringing anything up or they feel chesty but like … like the cough's like not really all the time …or, like so people always have confusing signs and symptoms. So, kind of being able to unpick that is quite helpful. [Participant 16 – Broad]*

> *"I think it's quite a good way of practising histories, I mean obviously practising histories on the wards is the best way of doing it. But if you can't be on the wards then it's quite a good way of doing it because, erm, doing it with your friends is a bit like, it's fine but…Yeah. It's not*

really the same and you don't really feel like you're testing yourself"
[Participant 10 – Focused]

Students who were 'Focused' and 'Thorough' felt that the way eCREST only provided information that was asked for and revealed information as the case went on, rather than all at once, was more realistic and allowed them to engage more with the scenario.

"I liked how information was released bit by bit. Because often when you're doing other things like on the case of the month, it's like the only timescale of the different… if the information being released is, as it's, as you read down the page. And it and, I like how it's you … because it's similar to what you do in a, in an actual consultation, you find out a bit of information and then that triggers you to ask more questions. And then you … a few minutes later you find out your history." [Participant 2 – Focused]

"the fact that you could order it, you could choose what information you got at what time, which is again more realistic than give me all the questions that you'd ask and then here's just a load of information which might not be related. So you only got the information that you asked for, which I liked. " [Participant 4 – Thorough]

8.4.4.2 Attitude towards learning using eCREST

Only one student in the study reported having a slightly negative attitude towards eCREST. This student had a 'Succinct' data gathering style. They expressed a lot of frustration at having to watch some videos of the patient and the video feedback from the GP; this was in part because they preferred text formats. However, it was also because they did not see the value of patient concerns and receiving feedback from eCREST. It seemed that this way of learning clinical reasoning skills was not engaging for this student. On the other hand another student who had 'Thorough' style noted that learning by doing,

even in a virtual setting, was helpful for them as it helped them to absorb the

information and attach meaning to the concepts in the cases.

> *"Well at first of all I didn't really care about their concern but then because everything was a no I thought that might be something that would give me a clear, so I guess it didn't in terms of me wanting to get more out of the patient." [Participant 5 - Succinct]*

> *"think I like it because I think I like learning whilst doing rather than like I would retain information a lot more if I, like when I attach patients to them. So like I often remember things because I've like seen it and gone oh yeah. And I had to like go through that process of thinking it through and then next time it's slightly easier." [Participant 11 – Thorough]*

8.4.4.3 Interactivity

Some students showed in the interviews that they felt eCREST was more

engaging than previously used online learning because it was more interactive,

in that students got specific responses to specific questions. Those who were

'Succinct' felt this was more realistic and more engaging compared to other

online learning that involved reading lengthy text and writing many responses,

as those tasks took a long time. However, one 'Focused' student did comment

that having the option to do some initial thinking using free-text responses may

have encouraged deeper thinking and would have been less led by eCREST.

> *"And the fact that it was interactive made it slightly more interesting as opposed to like. You know the case of the month that we do. There's a very read write, read write." [Participant 15 – Focused]*

> *"I like how they had the, erm, they had loads of questions that you could possible asked and then him saying the answer to all of them because it was kind of more interactive than when you have to type in questions and you don't get an answer, like on our case of the month." [Participant 5 – Succinct]*

8.4.4.4 Format

Some students noted that the cases were engaging because of the video format used to present the patient and the feedback. Students across all data gathering styles liked the videos, as they felt this helped students to feel more immersed in the simulation and feel like "someone's actually talking to you." Those who had a 'Focused' and 'Succinct' style also liked the transcripts, as some students commented that if they were to re-watch the cases it would be quicker to read the notes rather than watch the videos.

> *"I think it is definitely good that you have a transcript it would be really annoying have to play the video over and over again just to get like bits of information. So that was good. Yeah for sure. But I think, yeah I think it's good having like a patient there as well. " [Participant 11 – Thorough]*

> *"I personally don't like watching videos. So I thought it was nice that there was a transcript to allow me to read in my own time." [Participant 15 – Focused]*

8.4.4.5 Feedback

There were mixed views on the feedback provided to students. Some students across all data gathering styles felt it helped them to understand the reasons why their diagnoses were correct or incorrect. They also felt feedback was constructive and that the learning was useful. Others who were 'Focused' felt they needed more feedback, especially about their thought processes and what the essential questions to ask were. A few students across all groups also felt unclear what the correct diagnosis and management plan for each patient was. While eCREST was designed to show there could have been many different 'correct' diagnoses and different ways to manage the patient, some students perhaps needed guidance that is more specific.

"I think this would be much better than the resources that I had, when I was starting out just because it provides really detailed explanations for why this is possible and why this is not possible." [Participant 13 – Broad]

"Yeah, so if there was feedback on that process itself, 'cause if that's the main, if the emphasis is on the process, then feedback on the process rather than the end result would be useful. I think that's one of the, that's the USP of the programme is the process itself." [Participant 10 – Focused]

8.5 Discussion

This study used a mixed-method approach to explore how students reasoned while using eCREST and to explore what impacts eCREST and other factors may have had on clinical reasoning skills. The quantitative data gathered both in the feasibility trial and from the data collected by eCREST during the Think Aloud allowed me to characterise four different approaches to solving a patient case in eCREST. Students' approaches lay on two continua: being able to identify essential information to make an informed diagnosis (focus) and how much irrelevant information was gathered to identify that essential information (breadth). The four approaches identified were students being: 'Focused', 'Thorough', 'Succinct' and 'Broad'.

The qualitative data showed students were thinking about how focused and broad their data gathering technique was. They also often used learned data gathering strategies like the system-based approach and their limited clinical experience to help them test their hypotheses. Students' open-mindedness and flexibility about diagnosis varied. Students' data gathering, open-mindedness and flexibility in diagnosis were affected differently by the design of eCREST, with some students finding it provided structure and others

feeling it led them to ask random questions. Additionally, I identified several factors that were likely to have influenced clinical reasoning skills including: knowledge gaps, level of confidence and their engagement with the simulation.

8.5.1 Characterisation of different approaches to eCREST

An overview of the different data gathering styles and their associated characteristics is given in **Figure 8-1**.

Figure 8-1 Diagram showing different gathering styles and associated characteristics

302
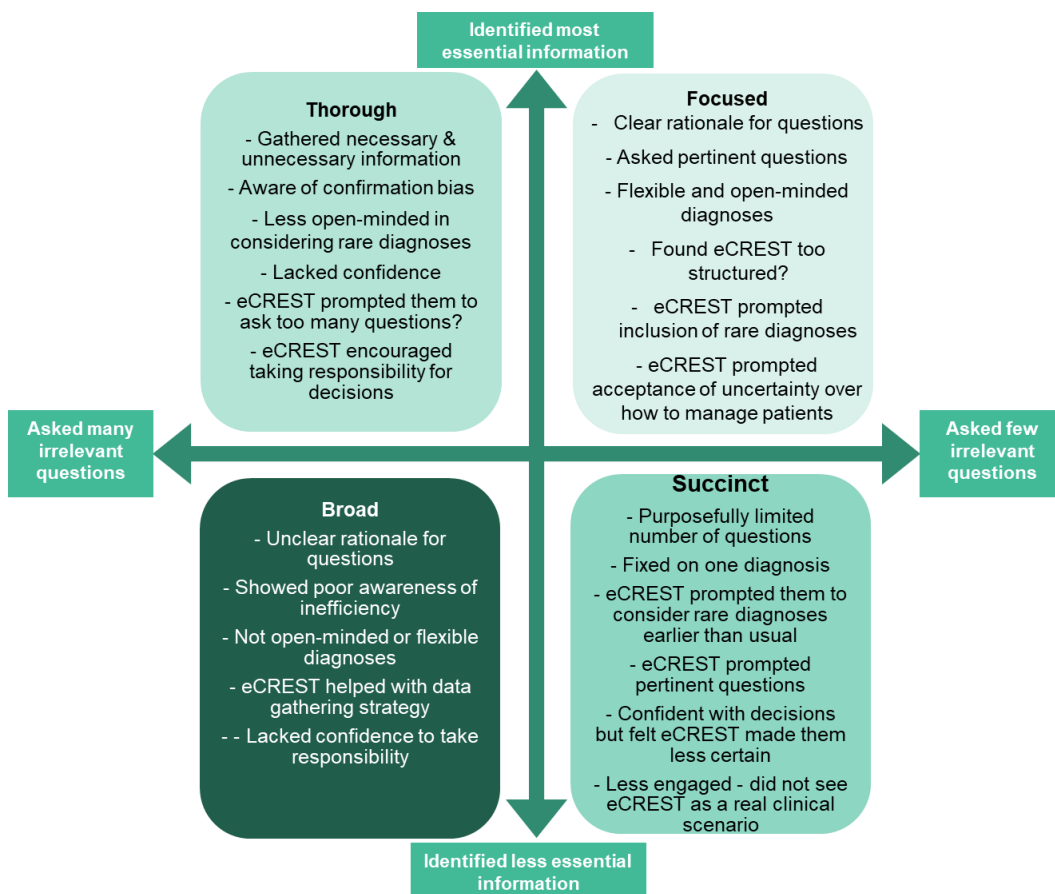
8.5.1.1 'Focused' style

These students gathered information by taking a system-based approach, learned from formal teaching and from clinical experience. They tended to have a clear rationale regarding the questions they asked and demonstrated that they limited themselves to only asking relevant questions. Students were relatively flexible and open-minded but showed some uncertainty about considering rare diagnoses. They found eCREST useful at times in guiding them to ask pertinent questions but also felt that it may have imposed too much of strict structure on data gathering. eCREST helped them to consider unlikely diagnoses earlier but they worried that in real practice it may be difficult and inefficient to investigate unlikely diagnoses. They recognised their lack of knowledge about managing patients in primary care. They felt they needed more confidence to make decisions but recognised that they needed to also accept this uncertainty and learn to manage their uncertainty through the safety netting of patients. They showed they were engaged with the simulation and found it useful to practise managing uncertainty in primary care. Those who used this approach were perhaps more independent thinkers and would have preferred more user input, such as using free-text responses before being prompted by eCREST.

8.5.1.2 'Thorough' style

These students used a predominantly system-based approach to gathering information and their goal was to be thorough and gather as much information as they could. They were aware of the influence of confirmation bias

during the task and tried not to seek information to confirm their hypotheses. However, consequently they gathered too much irrelevant information and asked many questions just to reassure themselves. These students sometimes found it difficult to keep an open-mind throughout the patient case, as they were too reassured by the absence of symptoms and tended to focus on one diagnosis. eCREST may have led them to ask more questions than they otherwise would have done. They misinterpreted some of the key red flag symptoms for lung cancer and recognised they had poor knowledge of managing patients in primary care. They lacked confidence in their ability to identify and manage serious diagnoses. They showed they feared getting the diagnosis wrong more than others, but they showed willingness to develop their skills by practising taking responsibility for making decisions, to improve their confidence in considering and investigating unlikely diagnoses.

8.5.1.3 'Succinct' style

Students gathered information based on learned strategies, by using system-based questioning and by basing questions on their usual clinical practice. These students approached the cases by limiting the amount of questions they asked. They showed awareness that cognitive biases influenced their clinical reasoning skills, but still struggled to consider and investigate alternative diagnoses. They struggled to stay open-minded, as they tended to focus on one or two diagnoses and they were too reassured by the absence of symptoms. They found eCREST useful in guiding them to ask pertinent questions and consider unlikely and serious diagnoses early in the consultation. These students felt that eCREST helped them to structure their information

gathering and be succinct but this was different from how they would usually ask questions. They had knowledge gaps of some of the diagnoses and showed they had misconceptions about the probability of diagnoses if symptoms are absent. They had poor knowledge of how to manage the patients but recognised this as a limitation. They were confident in their abilities but felt that eCREST made them feel more uncertain about their decisions. They saw eCREST as a game rather than an opportunity to practise their clinical skills as they would in practice, which may have led to a lack of engagement with eCREST. They liked that the multiple-choice options and lack of free-text made it quicker it complete a patient case.

8.5.1.4 'Broad' style

Students who used this style tended to ask many irrelevant questions and showed little awareness of the inefficiency of their strategy. They used eCREST to select questions that thought they would ask in real practice but showed little rationale behind their selection. They were aware that they were affected by cognitive biases, as they reflected that they generally sought information to confirm their hypotheses. They were not very open-minded about diagnosis and struggled to choose their diagnoses initially. They recognised the need to investigate unlikely diagnoses but questioned the efficiency of doing so. They were not very flexible with their diagnosis, as they did not change their diagnosis many times. They felt that eCREST helped them to visualise and organise their data gathering. eCREST did encourage these students to consider serious diagnoses early but they felt uncomfortable being led by eCREST to narrow down their focus and feared missing information. They

lacked confidence and were unsure of how they would balance the prioritisation and management of rare and likely diagnoses in reality. Students misinterpreted some information and had poor knowledge of management in primary care, which they acknowledged. They were somewhat engaged with the simulation but had too little confidence to feel comfortable with the responsibility of making decisions in the simulation.

8.5.2 Revisiting my theoretical model of students' clinical reasoning

My observation of undergraduate medical students' clinical reasoning skills during the Think Aloud gave me insight into how students reason, what other factors can influence reasoning and in what ways this can be exhibited in behaviour (data gathering strategies). The key findings from the Think Aloud were added to my theoretical model of how students reason and how online patient simulations might influence reasoning (highlighted in **Figure 8-2**). The results of this study also have implications for the OPS logic model of how online patient simulations can help students learn clinical reasoning skills but this will be discussed in more detail in **Section 9.3.1.1**.

Figure 8-2 Updated theoretical model of learning clinical reasoning skills
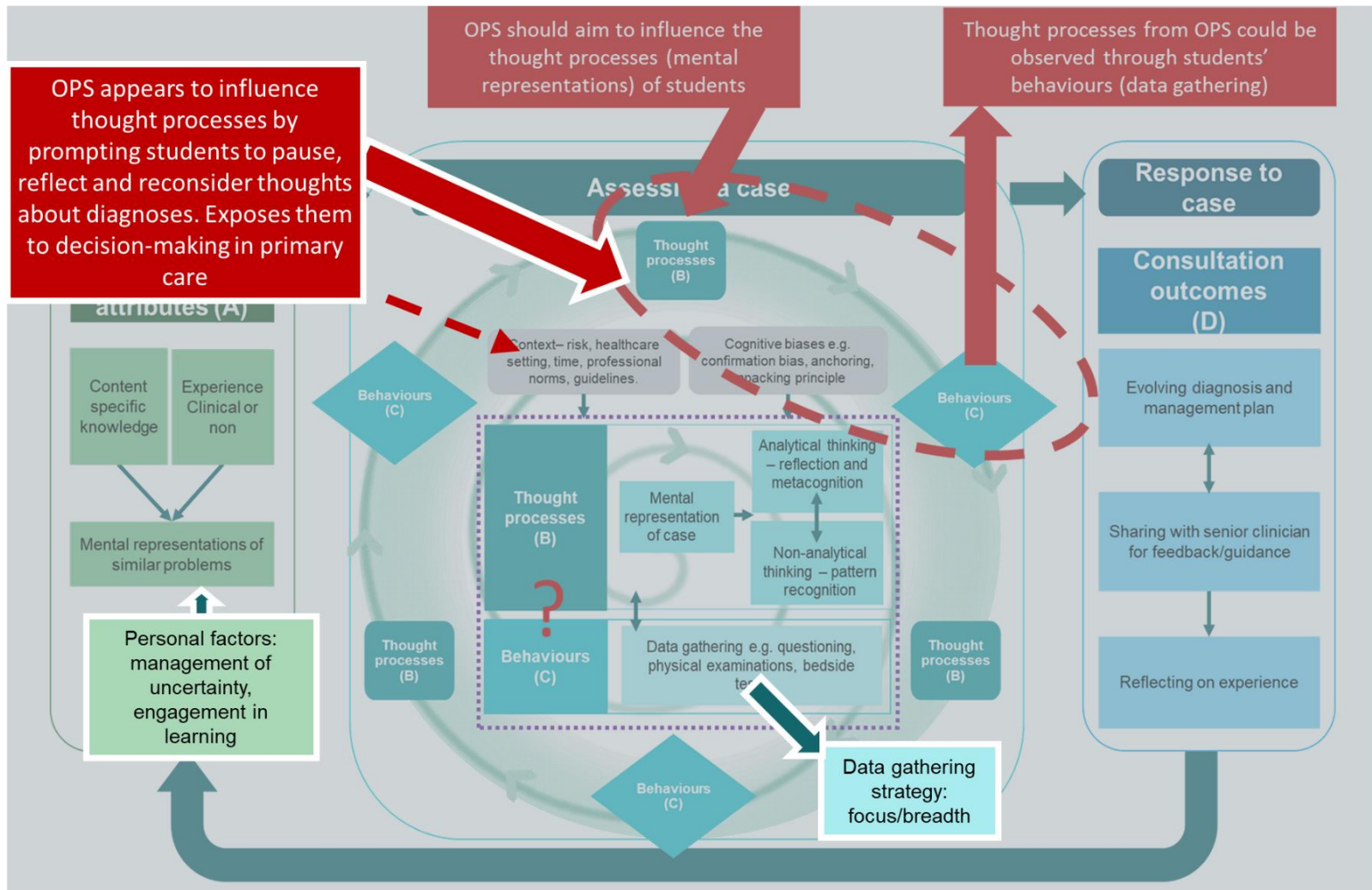
Personal factors, in particular feelings of confidence or self-efficacy, appeared to affect students' thinking and ultimately how they gathered information. Therefore, I added this factor to my theoretical model as a factor prior to their current case (Phase A) that influences the student's initial impressions of the current patient presentation (**Figure 8-2**). I found those who were less confident appeared to ask more unnecessary questions (those who were 'Thorough' and 'Broad'). An assumption could be made that other personal factors such as personality, which may be related to confidence, could also affect reasoning. Furthermore, how engaged students were and their attitudes towards eLearning could have influenced reasoning. Those that were less engaged in learning skills may have engaged in less analytical reasoning, as this is more effortful and slower than non-analytical reasoning; this may explain why those who used a 'Succinct' approach asked fewer questions. However, given students may not have adequate mental representations of problems to be able to recognise all the essential information of a case this could be a risky strategy in practice.

Another finding from the Think Aloud that influences the theoretical model was that students demonstrated their ability to test their diagnostic hypotheses through data gathering via selection of data gathering styles. These data gathering styles are measurable from the data collected by eCREST, such as the proportion of essential questions they asked out of the number of essential questions available. I added data gathering strategies into my theoretical model in Phase C, as observable behaviours related to clinical reasoning (**Figure 8-2**). The characteristics of these styles were developed

further through analysing the student's verbal reports of their thinking in the Think Aloud and interview study. These styles, as described above, fell across two continua: focus and breadth. This could reflect the use of an interaction between the different types of reasoning: analytical and non-analytical reasoning. For example, those that were more 'Focused' might have used more non-analytical reasoning, as they were able to identify the most relevant and essential from the patient, without wasting time on unnecessary questions. Those who took a more broad patient history may have engaged more analytical thinking strategies, and gathered a lot of information to confirm and refute their many hypotheses. It was likely that all students used both strategies at different points in the case, perhaps using analytical reasoning when an unexpected symptom arose. It is important to recognise that the strategies identified in this study are unlikely to be exhaustive and definitive. Further investigation into the different data gathering strategies of students could help medical educators understand what aspects of students' reasoning could be improved and how and online patient simulation could be designed to specifically address this.

I also observed in the Think Aloud that online patient simulations did seem to have a positive influence on reasoning, as students reported that eCREST helped to: guide their data gathering to be relevant and focused on their diagnostic hypotheses through the prompts in the system; be flexible and open-minded in their diagnoses and gave them time to reflect on their reasoning (see red box in **Figure 8-2**). These influences seemed to alleviate some of the influence of the cognitive biases that can affect reasoning. Therefore, the design

of eCREST, i.e. the provision of prompts to revise diagnoses and lists of questions in eCREST, were largely successful in influencing thought processes, which was reflected in the students' behaviours (data gathering styles). However, for some students the design may not have been helpful. Some students mentioned selecting questions randomly without thinking about why they were asking them or selecting all the questions because they were available. This behaviour could be an artefact of the simulation environment, as in real life interactions with patients students may be less likely to ask questions randomly or to be thorough due to time pressures and pressure of being observed by a senior clinician. While it may suggest students are not thinking or engaging with the simulation for that particular case it could also be that they are employing a learning strategy of trial and error. They may be using the first case to explore the information they can get from the simulation and testing how focused and broad their strategy needs to be to reach an appropriate diagnosis. This strategy of trial and error through repeated practice is recommended by learning theories such as deliberate practice (Ericsson, 2008; Ericsson et al., 1993). As I only observed students completing their first case this hypothesis would need to be tested in future research. Further consultation with medical educators could help to establish whether prompting students to ask questions and reconsider their diagnoses is an acceptable approach to teaching reasoning to students or whether eCREST could be adapted to avoid students randomly selecting questions.

Another benefit of the simulation was that it exposed students to clinical reasoning in uncertain situations in primary care (see dashed red arrow in

Figure 8-2 to demonstrate the influence of eCREST on context). I found students demonstrated they were uncertain about diagnosis and management of the patient in the consultation. Uncertainty regarding diagnosis and data gathering increased as the case developed, as the patient's symptoms were vague and non-specific and the absence of red flag symptoms was revealed. These scenarios are common in primary care but students seemed unsure as to the best strategy to reach an appropriate diagnosis in these cases. The main confusion was whether they should include and investigate possible but rare diagnoses or whether they should focus on the most likely diagnoses. Managing patients in primary care was also an issue because students were unsure of how to appropriately safety net patients in primary, which was likely due to their limited clinical experience in these contexts and lack of knowledge of appropriate guidelines (Harding et al., 2015; McDonald et al., 2016). Previous research has also demonstrated that managing ongoing uncertainty is a distinct phase in the clinical reasoning process even for clinicians. For students and junior doctors who less able than senior clinicians to tolerate and accept uncertainty this can cause a great deal of anxiety (Adams, Goyder, Heneghan, Brand, & Ajjawi, 2017; Mamede et al., 2008). Through interviews with junior doctors previous studies have formulated recommended strategies for managing ongoing uncertainty, such as sharing the responsibility of decision-making, safety netting and considering the worst case scenario, which could help to inform feedback provided to students on how to manage and accept uncertainty (Adams et al., 2017). Completing repeated cases in eCREST and receiving feedback on managing uncertainty in these contexts could help students to develop better strategies on how to manage uncertainty and lead

them to the appropriate guidelines. Furthermore, improving these skills could help to reduce the amount of missed diagnostic opportunities in primary care, which could contribute towards better outcomes for patients (Rubin et al., 2015).

8.5.3 Comparisons with other literature

Previous studies using the Think Aloud approach to study students' reasoning skills found that students were more likely to follow confirmation lines of reasoning and consider only expected findings for a given disease than clinicians (Johnson et al., 1982). I also found that students were focused on seeking expected findings for a disease and were too reassured by the absence of symptoms. A further Think Aloud study also found that beginner students tended to ignore or reinterpret inconsistent evidence to fit their hypothesis (Arocha & Patel, 1995). Similar to these studies, I found most students were likely to be subject to cognitive biases, in particular confirmation bias. They were fixated on one diagnosis, they asked confirmatory questions and did not always explore alternative explanations for the symptoms. Some showed awareness of this during and after the Think Aloud task but this did not always seem to help them with their thinking. Indeed, one student who verbalised that they were trying to ask non-confirmatory questions also asked many unnecessary questions. Several studies have shown that just being aware of cognitive biases does not appear to improve clinical decisions (Monteiro & Norman, 2013; Norman et al., 2014). It is important to recognise that confirmation bias affects even expert clinicians and does not necessarily lead to errors but students may be in need of more guidance on how to overcome this bias in certain contexts like primary care where it may lead to serious missed diagnostic opportunities

(Berner & Graber, 2008). It is possible that further practice with eCREST could help students to develop strategies when appropriate that avoid confirmation bias, as it encourages students to consistently consider alternatives and justify decisions.

8.5.4   Strengths and limitations of this study

A strength of using the Think Aloud approach to capture clinical reasoning skills was that it provided rich insight into the thought processes of students as they used eCREST in real-time. Consequently, there was less bias in students' reports of their thought processes compared to a retrospective report, which are more susceptible to influence from other information in their long-term memory. The Think Aloud also helped to identify how students reasoned when using eCREST and how eCREST worked as an intervention, which added to the theoretical and logic model used in this thesis. The semi-structured interviews that followed the Think Aloud also allowed students to elaborate on their thought processes and further reflect on their abilities and eCREST.

A strength of the study was also that it took a mixed-methods approach, which meant my analysis of the qualitative data informed the quantitative and vice versa. The use of two different methods allowed for different kinds of research questions that were important to this evaluation to be answered. The Think Aloud provided insights into how students actually reasoned when using eCREST and gave rich detailed data on the reasons why students were gathering data in certain ways and the other thoughts that influence these

decisions. However, using mixed-methods did have some limitations, as I was the sole researcher and was working within the timeframe of a PhD. Therefore, there was limited time and resources to collect and analyse both types of data.

There were some limitations to using the Think Aloud approach. One of the main criticisms of Think Aloud studies regards the validity of participants' responses given they are being observed and prompted when they become silent (Charters, 2003; Cotton & Gresty, 2006). It is likely that students will attune their responses due to social desirability. For example, in the case of the medical students they might have felt they needed to overstate their confidence or knowledge in an attempt to appear more knowledgeable (Cotton & Gresty, 2006). However, I tried to mitigate this effect by reassuring participants that the purpose was not to evaluate their performance on the case but to explore how eCREST helped or hindered them in this process. Additionally, I found many students referred to feeling unconfident and acknowledged a lack of knowledge, which suggested to me that most were comfortable expressing their thoughts related to their skills and eCREST and my presence did not appear to hinder them from speaking their mind. It is also likely that the prompts could have changed students' thought processes. For example, it is possible they were more reflective than they would have been if unprompted or unobserved. Therefore, caution should be taken when interpreting results, in terms of the extent to which eCREST itself influenced reflection, rather than the Think Aloud approach. .

To attempt to minimise my influence on students' thought processes, most of my prompts were generic but I needed to use some more directive prompts, such as "why did you chose that?", to help participants understand what types of thoughts to think aloud. Previous studies have shown that generic prompts do not influence students' thoughts but the more directive prompts may have changed students' thought processes(Ericsson & Simon, 1980). However, I found that using these prompts provided richer data on students reasoning during eCREST, as when practising the method I found students were likely to remain silent or feel confused when only generic prompts were used. Nevertheless, future research could consider other ways of exploring students' thought processes when using eCREST, to minimise the impact of the researcher on the data. For example, students could retrospectively recount their thought processes after using eCREST. This data could be used in conjunction with Think Aloud data on students' (prompted) concurrent thoughts, to triangulate the data and build a greater understanding of how students reason when using eCREST.

Another limitation of the method is the complexity in the interpretation of Think Aloud data. Some utterances from participants can be ambiguous and rely on the researcher to infer meaning (Charters, 2003; Cotton & Gresty, 2006). I attempted to overcome this by having a clinician read over all of the transcripts and provide feedback on what they inferred from the students verbal reports to provide a second and alternative perspective. I also followed-up on any ambiguous comments students made by prompting them to explain their thoughts during the Think Aloud and follow-up interview. Some students were

also better at thinking aloud than others were. Therefore, it was difficult to distinguish whether students who did not verbalise much in the Think Aloud had limited reasoning, were disengaged from the task or were just uncomfortable thinking aloud.

How I selected students to take part in the Think Aloud might also be a limitation of this study. I used convenience sampling and snowballing to recruit students. This could have potentially biased the data and potentially only represented the thoughts of students who were highly engaged in eLearning or research, rather than the average medical student. However, my Think Aloud analysis showed that some students demonstrated a lack of engagement with eCREST and provided valuable insight into how students with low engagement would respond to eCREST. The use of monetary incentives could have helped to ensure that students who were generally less engaged with research, eLearning and learning outside of their curriculum were represented, as their incentive to take part was financial. However, equally this could also have introduced a selection bias. Purposeful sampling by selecting students who fell into the extremes of the continua or demonstrate typical behaviours for each data gathering strategy could have helped to define the different styles more clearly and future research in this area may benefit from this approach.

8.5.5 Summary

This chapter aimed to address research questions four and five of my thesis, which were to explore how students reason and the factors that influence reasoning. From quantitative and qualitative data I found that students had

different data gathering strategies that lay across two continua: how focused they were and how broad they were with their questioning. The design of eCREST was helpful for most students by providing them with a structured and guided way of taking a patient history and helping them to be more flexible and open-minded about diagnoses. However, for some it may have encouraged them to ask many more questions than were necessary or random questions. There was evidence that a lack of confidence in students' reasoning skills but exposing students to cases with a large degree of uncertainty in primary care could help students to come to accept and manage their uncertainty more effectively. A lack of knowledge and engagement in learning through the simulation also was found to affect the reasoning students. However, these factors could be overcome if the simulation was integrated into an appropriate module, where they would also gain the relevant knowledge to solve the case and feel more motivated to engage with the simulation, as it would be more relevant for their current learning. Further research is needed to explore these data gathering strategies in more depth, possibly by selecting and observing students with extreme behaviour and typical behaviours.

There were several indications for how the design of eCREST could be improved, such as by providing more support to help students gain confidence in making decisions and by providing clearer indications of how they should have approached the patient cases. This chapter also helped to inform my understanding of how students reason, allowing me to revise my theoretical model proposed in **Chapter 1**. In **Chapter 9,** I will demonstrate how these findings have helped to refine the logic model for how online patient simulations

can improve students' reasoning. In **Chapter 9,** I will also use the qualitative

data to explore why students in the intervention group of the feasibility trial

appeared to take a more 'Thorough' approach to reasoning than the control

group.

9 Discussion

9.1 Summary of findings

In this thesis, I describe my contribution to the development and the evaluation of an online patient simulation tool (eCREST) designed to improve undergraduate medical students' clinical reasoning skills. It advances the current literature by proposing a logic model for how online patient simulations can improve clinical reasoning and by developing novel ways of identifying and assessing clinical reasoning skills.

The research questions of this thesis were addressed by firstly undertaking a systematic review and meta-analysis **(Chapter 3)** that identified gaps in the current understanding of online patient simulations and established the need for methodologically robust evaluations of online patient simulations (RQ 1). I then proposed the OPS logic model of how medical students' clinical reasoning skills can be promoted using online patient simulations based on theory and empirical evidence (RQ 2, **Chapter 4**) and how I contributed to the development of an online patient simulation learning tool (eCREST). I assessed the feasibility and acceptability of eCREST at three UK medical schools through a feasibility study using a RCT design **(Chapters 5 & 6)**. Results indicated that eCREST was feasible and acceptable, and warrants a full-scale definitive RCT to estimate effectiveness of such tools on reasoning. Furthermore, I developed a measure of clinical reasoning skills, based on the key features of problem, which suggested that eCREST could potentially improve how students gather information from a patient (RQ 3). To explore how students reasoned when

using eCREST and the factors that influenced reasoning I used quantitative data collected from eCREST along with qualitative data from the Think Aloud and interview study **(Chapters 7 & 8)**. This allowed me to test the OPS logic model and refine my theoretical model of how students reason. I identified several different ways that students approached data gathering and how the design of eCREST, student confidence and their engagement with eCREST influenced these different strategies (RQ 4 & 5).

9.2 Strengths and limitations of my thesis

9.2.1 Intervention

A strength of eCREST was that it provides a new model of how to teach clinical reasoning skills, by prompting and guiding students to reconsider their diagnoses at the beginning and throughout a consultation and providing feedback. This model was based on the OPS logic model, developed as part of this thesis and informed by theories of learning and clinical reasoning. There are ways in which eCREST as an intervention could be improved. Students' comments on the acceptability survey from the feasibility RCT, and in the Think Aloud, provided insight into several areas of potential improvement if eCREST were to be developed further. Some of the user comments aligned with what theories of learning and cognition suggest would help students to learn skills. For example, users reported they would like more cases to practise with, which according to deliberate practice would help students to learn by trial and error (Ericsson, 2008). Theories of cognition also suggest that having multiple cases can improve cognitive processes by providing them with a greater number of exemplar cases, which will help to expand their mental representation of those

problems (Cook & Triola, 2009). Users also suggested they needed more feedback on their reasoning and the essential questions they should have asked. The literature shows that feedback is a key aspect experiential learning and can enable reflection (Kolb, 1984; Mamede et al., 2008). Therefore, clearer feedback would be an important aspect of eCREST to improve but caution would need to be taken in how this was presented, so as not to suggest there was only one correct diagnosis and way of approaching the case. Students had mixed views on how much user input (via free-text responses to the questions) should be provided. While it could be argued that less user input could help to reduce cognitive load and reduce user fatigue, it could be useful to include some free-text responses at key stages in students' thinking, such as when the student initially generates hypotheses, to allow students to compare their initial thoughts to those suggested by eCREST. Further user testing will help establish the appropriate amount of free-text to ensure eCREST remains user friendly.

9.2.2 Mixed-methods

The strengths of my research were that it used a mixed-methods research design. This allowed me to answer a broader range of research questions (Creswell et al., 2003; Johnson & Onwuegbuzie, 2004). It allowed me to use quantitative research methods to answer questions regarding uptake and retention, acceptability and potential impacts of eCREST on clinical reasoning skills. I used these quantitative findings to inform qualitative questions including how students reasoned during eCREST, how other factors affected reasoning and how the OPS logic model performed. Furthermore, using a mixed-methods approach combined the strengths from both research methods (Creswell et al.,

2003; Johnson & Onwuegbuzie, 2004). The feasibility RCT provided less biased and more generalisable findings on clinical reasoning and potential impacts of eCREST, while the Think Aloud and interview data provided in-depth data on clinical reasoning and how eCREST works as a learning tool.

There were disadvantages of using the mixed-methods approach as it was practically difficult to conduct the research as one researcher. Due to the time constraints of my PhD I had to carry out the data collection of both methods concurrently instead of following a purely sequential mixed-methods design as intended. This may have led to some biases in the way I collected and analysed the data from both research methods, as my preliminary knowledge of the data might have guided me unconsciously in a particular direction when analysing the data. Moreover, conducting a simple purely sequential mixed-methods study is not necessarily realistic, as the process of collecting and analysing two types of data is complex and was more iterative in nature than sequential. In practice I found that qualitative data analysis could help to explain and interpret the quantitative data but also the quantitative data could help to organise the qualitative data through the use of typologies. Describing the process of how the qualitative and quantitative data informed each other at different stages of the analysis was challenging. For the purposes of improving the rigour and clarity of my mixed-methods approach I amended my initial mixed-methods process diagram (see **Section 7.10**) and described in detail how one type of data informed the other throughout my PhD.

9.2.3 Selection bias

Selection bias in the feasibility RCT has implications on the generalisability of the results, as it is likely that only students who were already engaged with eLearning and who regularly participate in extracurricular learning took part. One way to resolve this issue would have been to integrate eCREST into the curriculum and ensure every student was required to complete it. However, a newly developed and untested online learning tool like eCREST is unlikely to be integrated into an already saturated curriculum. My feasibility RCT established the proof of concept for using eCREST in UK medical schools, so the next step for research on eCREST would be to attempt to integrate eCREST into medical curricula and conduct a full-scale definitive RCT, which would avoid selection bias and have fewer issues with uptake and retention.

In the Think Aloud study, a more purposeful sampling strategy could have been undertaken to reduce selection bias. The sample I chose was based on convenience and from snowballing methods, which meant that I may not have sampled representatively across the data gathering strategies I identified. A better sampling strategy might have been to select a sample of students from the feasibility trial that exhibited typical or extreme quantitative data scores for each of the strategies I identified. This would have helped to inform the different strategies students use in online patient simulations and why.

9.3 Key learnings from my PhD

There were three key findings from my PhD that have wider implications for future research in this area. This synthesis of my findings addresses my final research question six and seven, which was to recommend how online patient simulations should be designed and evaluated. The three main findings were:

1. It was feasible to evaluate online patient simulations in medical education using an RCT design. However, there were significant challenges to doing this because of the lack of current evidence of their effectiveness and difficulties in implementing teaching innovations into the curriculum;

2. It was possible to assess clinical reasoning skills using validated approaches. Assessment of clinical reasoning skills was difficult because of the complex nature of clinical reasoning but the use of measures, such as the key feature problems, provided a way to assess the impact of an intervention on reasoning;

3. The effect of online patient simulations on reasoning may be influenced by students' data gathering strategies. Online patient simulations helped most students to structure how they gathered information from a patient and helped them to maintain flexibility and open-mindedness when considering their diagnoses. However, factors such as a student's confidence in their skills, and their lack of knowledge and engagement in learning also influenced their reasoning.

9.3.1 The feasibility of developing and evaluating teaching innovations in medical school settings

9.3.1.1 Developing digital innovations

I found it was possible to develop a theoretical and evidence based online patient simulation that could be integrated into an undergraduate medical school curriculum. While it may take an initial investment of time and resources, the benefits of developing an adaptable and updatable learning resource that could be used for years to come may make this a worthy investment. I found a wealth of theories from psychology and education to support the use of online patient simulations to teach and improve reasoning skills and provide insight into the required features for such tools. However, there was a lack of empirical evidence to support the use of such tools to improve reasoning. My systematic review found only 12 comparable previous studies that had evaluated online patient simulation tools for the purpose of training clinical reasoning skills in students, none of which provided detailed descriptions of how they were developed and many lacked methodological rigour. With the current evidence, it is not yet possible to identify what intervention features are most effective in teaching students' reasoning or to test different tools against each other. Yet theories of learning and clinical reasoning and evidence from current teaching suggest that features such as reflection, scaffolding and feedback are useful features to include in online patient simulations (Posel et al., 2015). How these features are delivered, however, does vary, as shown in my systematic review. Thus, further research is needed to determine the most effective way of providing such features.

As noted in my systematic review, few previous evaluations have provided detailed descriptions of how they used theory to inform their interventions or provided detail on how they piloted and developed such tools. I developed a logic model to show how clinical reasoning skills can be improved by online patient simulations (described in **Chapter 4**). I updated this model based on my analysis of the qualitative data from the Think Aloud and interview study. This updated logic model could help future developers and evaluators of online patient simulations to design and evaluate other theoretically and empirically informed learning tools. **Figure 9-1** shows the logic model with the added moderators (numbered 1 in diagram), mechanisms of change (numbered 2 in diagram) and measurable outcomes (numbered 3 in diagram).

My qualitative data highlighted moderating factors that could influence students' reasoning during eCREST that I had not previously included. Personal factors including students' confidence and uncertainty in managing patients in primary care, personality, and how well they engaged with the simulation and learning, all influenced students' reasoning - this was demonstrated in their behaviours (i.e. the questions they asked the patient and their reported reasons for asking those questions). The design of the online patient simulation was also another moderating factor that influenced reasoning. I found in my qualitative study that most students found that eCREST helped them to structure how they gathered their data and to be more flexible and open-minded about diagnoses. However, for some students the way eCREST was designed may have led them to ask random questions and encouraged passive learning. Understanding the role of these moderating factors can help improve our understanding of

students' reasoning when using online simulations and what further support

could be provided by eCREST to improve expected outcomes.

Figure 9-1 Updated logic model based on qualitative findings
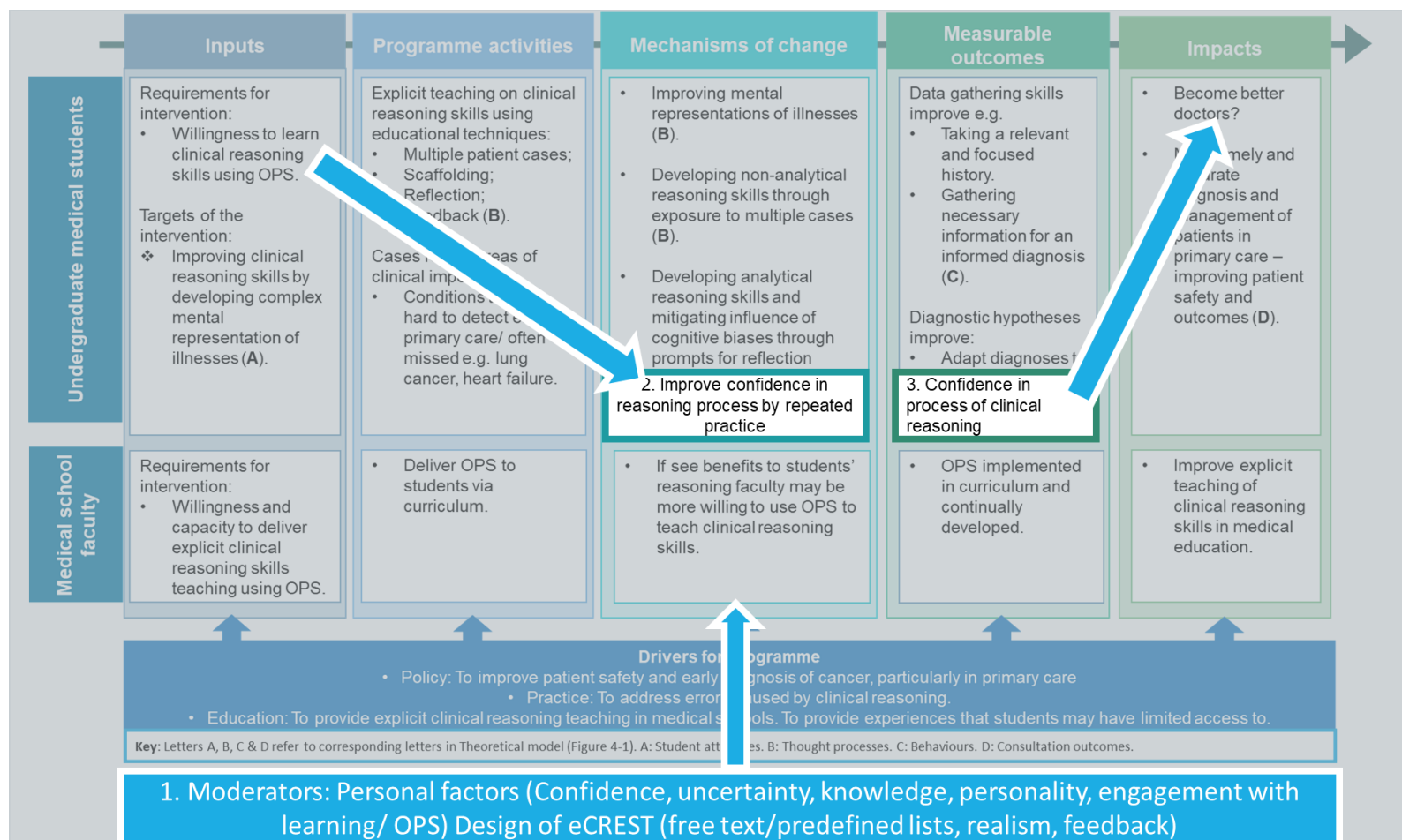
I also included improving confidence through repeated practice as a mechanism of change (numbered 2 in **Figure 9-1**). I found in the Think Aloud and interviews that some students felt that online patient simulations could help them to improve their confidence and self-efficacy, particularly in dealing with uncertainty in primary care. The feedback and repeated practice provided by online patient simulations can help students to understand their strengths and weaknesses, which can help them to gain an appropriate level of confidence. The confidence in skills that could potentially be gained by using online patient simulations could improve students' willingness to use online patient simulations, which would be required if the tool were to be tested (as indicated by the left arrow in **Figure 9-1**). Developing students' confidence in their clinical reasoning skills could help students to make decisions more efficiently and feel more motivated to learn and develop their skills further (Bandura, 2001; Burke & Mancuso, 2012). Confidence in clinical reasoning skills is measurable through a student's perception of their confidence in their skills (numbered 3 in **Figure 9-1**). The impacts of improving confidence in clinical reasoning could contribute towards students becoming better future doctors, which is one of the potential impacts of eCREST (as indicated by the right arrow in **Figure 9-1**).

9.3.1.2 Conducting evaluations in medical school settings

I found that it was feasible to conduct an RCT to test online patient simulations in multiple medical schools. However, evaluating digital innovations in medical school settings was challenging because it was difficult to integrate them into the curriculum and test them in the way they are intended to be used in practice. Consequently, the feasibility RCT I conducted, using student

329

volunteers, does not provide information about the feasibility of using eCREST within the medical school curriculum. It was likely that I found it difficult to integrate a newly developed online patient simulation into curricula because it had not already been tested and there was no evidence of proof of concept. Furthermore, it is difficult to introduce new content into in an already saturated curriculum. In my feasibility RCT I was able to somewhat integrate eCREST into the curriculum at SA (for cohort 2), as students participated when on their GP placement module, although I still relied on student volunteers. This improved uptake compared to advertising eCREST to students on their teaching days and via newsletter or email, which were not specifically linked to any relevant modules. Uptake likely improved because students saw the training as relevant to their module; therefore, they perceived the training to be more useful and had more motivation to take part. Moreover, eCREST would have seemed more connected to the medical school itself, rather than an externally imposed tool, which may have given the students more confidence that eCREST was a learning tool that was reliable and helpful. Key factors relating to recruitment failures, which may also be common to other online resources, were lack of information about the resource and technical problems. However, these factors could be overcome with rigorous user testing and discussions with users on what information they need to know about online resources before using them. Further evaluations of online patient simulations should be integrated into a specific module in the curriculum that is relevant to the content, to improve engagement from medical schools and students and to make the learning more relevant for the students.

Overall, I found three medical schools in the UK were willing to test an online patient simulation designed to help improve students' reasoning. However, medical schools had some reluctance about testing eCREST using a RCT study design. This raises concerns over the willingness of medical schools to participate in a definitive full-scale RCT to test the effectiveness of eCREST, which is the recommended approach suggested from my feasibility RCT. The major concerns were ethical concerns about the fairness of providing eCREST to only students in the intervention group. All students were provided with eCREST at the end of the study but there were some concerns it may unfairly advantage students who had earlier access. RCTs are useful study designs, as it is possible to estimate the cause and effect of an intervention on outcomes and they reduce the chance of selection bias (Campbell et al., 2000). Given the ethical concerns of medical schools, it may be that a specific kind of RCT, a randomised crossover design, might have better uptake from medical faculty, as it would involve exposing both groups to an intervention but at different periods (Sibbald & Roberts, 1998; Sullivan, 2011). Alternatively, other types of study design may be more practical for further evaluation, as they may have less ethical issues, such as qualitative approaches. However, these study designs may not provide data on the effectiveness of online patient simulations and my systematic review showed that there is a need for more robust effectiveness studies to progress the development and understanding of online patient simulations.

9.3.2 Assessing clinical reasoning skills

9.3.2.1 Defining clinical reasoning skills

My literature and systematic review highlighted the complexity of measuring clinical reasoning skills (Higgs et al., 2008). One of the reasons it is difficult to measure these skills is because the definition of clinical reasoning skills is broad and covers a range of skills and behaviours, including gathering and interpreting data. It was unclear before conducting my research what specific aspects of clinical reasoning skills and their associated behaviours that online patient simulations would most benefit. My theoretical model of students' clinical reasoning skills described in **Chapter 1** shows how I defined and conceptualised clinical reasoning skills and my logic model described in **Chapter 4** began to map out how online patient simulations could benefit reasoning. Data from my quantitative and qualitative studies indicated that the main aspects of clinical reasoning skills that online patient simulations impacted on were how students reached an appropriate diagnosis rather than how accurately they can diagnose patients, which somewhat corroborated with my logic model. I did expect that the online patient simulation might improve consultation outcomes like diagnostic accuracy and not just data gathering skills. However, it was likely the educational techniques used in the simulation were targeted at changing students' thought processes rather than their knowledge of symptoms and diagnoses, which may have more of a bearing on diagnostic accuracy than on thought processes. It may be that if online patient simulations are fully integrated into a relevant module then students would gain complementary learning through other methods to improve their knowledge, which would in turn improve their diagnostic accuracy. This information can help

medical educators to understand the role of online patient simulations in the teaching of clinical reasoning skills and what aspects of clinical reasoning skills that they may be less adept at training, such as knowledge of diagnoses and management of patients. Future researchers can also have a clearer understanding of what changes to clinical reasoning skills to expect from online patient simulations and be able to more precisely define and, therefore, observe and measure these outcomes.

Based on my research a definition of clinical reasoning skills for medical students in the context of primary care is: the use of one or more data gathering strategies that enables students to take a focused and relevant patient history and identify the necessary information to make an informed diagnosis. It requires the ability to be open-minded about rare diagnoses and show flexibility in refining that diagnosis based on the information they have gathered. Clinical reasoning skills at this level of education and in the primary context are limited by a student's lack of knowledge of related symptoms and diagnoses and management of patients in primary care, and by their confidence and ability to accept and manage uncertainty. **Chapter 8** shows how this can be graphically represented in the revised theoretical model of students' clinical reasoning skills.

9.3.2.2 *Validity and reliability of measures of clinical reasoning*

I found that it was possible to assess clinical reasoning skills using online patient simulations. I developed a clinical reasoning measure, based on the data automatically collected on a further patient case. This method was

appropriate as it ensured that students were assessed on clinical reasoning skills in the same context as to what they were taught in the online patient simulation intervention. Indeed, my systematic review found many studies also took this approach. I also found it useful to use a previously validated measure of clinical reasoning (the key features problem) as a framework for how to assess clinical reasoning skills via a patient case (Page et al., 1995). Ultimately, clinical reasoning skills are difficult to measure because it is not a generic skill but is highly dependent on knowledge, previous clinical experience and clinical context (Higgs et al., 2008). Therefore, there is no generic measure of clinical reasoning skills available. Using a previously validated framework to construct a measure of clinical reasoning skills is perhaps the closest researchers can get to a 'generic' measure of reasoning that could be compared across studies to determine effectiveness.

I showed it was possible to measure some aspects of the validity and reliability of clinical reasoning measures. My systematic review indicated that reviewed studies were poor at demonstrating the validity and reliability of their outcome measures. I validated my measure by checking with a panel of clinicians and my supervisors that the measure had content and face validity. I also sought criterion validity, by exploring whether my key features predicted key consultation outcomes, such as identifying the most serious diagnosis. I established reliability by using my measure to assess reasoning across all the patient cases. I found mixed results of the validity and reliability of my measure of clinical reasoning skills, so more research is needed to identify whether the key features I identified could be used in future studies. It is challenging to

establish the validity and reliability of such measures given that evaluation studies in medical education tend to have small sample sizes and it is generally not their focus. However, future researchers need to focus on assessing the psychometric properties of their measures and use previously validated measures, such as the key features problem, at least as a framework to evaluate the effectiveness of their interventions. If I were to conduct my research again I would consider creating the key features problems using the exact procedure outlined by Page et al. (1995). Indeed, my key features measure may have been less valid and reliable than a true key features problem because I did not follow the exact process (Bordage et al., 1995; Farmer & Page, 2005; Page et al., 1995). For example, I did not create new patient cases and create questions for students based on that case; instead, I used a patient case already developed in eCREST and data already collected from the cases. I did this because it would have been very resource intensive to develop new cases and questions for the key features problems (Farmer & Page, 2005). It was more practical, given the time constraints of my PhD, to use an already developed patient case in eCREST; particularly because the patient cases in eCREST gathered rich data on the real-time clinical reasoning skills of students at critical steps in the process of making a diagnosis, much like the key features problems. More research, testing the psychometric properties of my approach, is needed to determine whether this approach to developing key features problems is valid and reliable.

9.3.2.3 Alternative measures

I established that there were several ways to assess clinical reasoning. In addition to the key features measure of clinical reasoning, I used a validated sub-scale of a self-reported survey that measured clinical reasoning, (the FIT survey) (Bordage et al., 1990). I chose this measure because it evaluates the reasoning processes behind making clinical decisions and was practical to administer online (Bordage et al., 1990; Higgs et al., 2008; Round, 1999). However, I found that this measure may not be sensitive enough at detecting small changes in clinical reasoning skills. Furthermore, the validity of this measure was questionable as it was a self-report measure and research has shown that students commonly overestimate their skills and have a poor self-perception of performance (Dunning, Heath, & Suls, 2004). My Think Aloud and interview study also indicated that medical students have a poor perception of their performance. Self-report measures require students to reflect upon their skills and can be biased by their previous knowledge and experiences. The benefit of using a patient case and developing a key features problem is that it can capture actual performance and provides a potentially less biased indication of their actual skill. Therefore, I would not recommend the use of the FIT to measure clinical reasoning skills in future evaluations over using patient cases. However, I would argue that it is still important to capture students' perception of learning but this could be captured using a much shorter survey, such as the acceptability survey used in my feasibility RCT.

9.3.3 Online patient simulations may influence data gathering strategies

The final key finding from my PhD was the deeper understanding of the different styles of gathering data that students use as part of their clinical reasoning skills. Both my quantitative and qualitative data revealed that students use different approaches when gathering information from a patient. Data gathering strategies appeared to lay across two continua: how focused students were at identifying questions to make an informed diagnosis and how broad their questions were. Furthermore, I found that factors such as the design of eCREST, confidence and level of knowledge might contribute towards the selection of different strategies. However, there remain many unanswered research questions, such as what conditions or contexts lead students to use these different strategies or change strategies.

My qualitative data supported parts of my logic model outlined in **Chapter 4**, as I found that online patient simulations did help to facilitate analytical thinking through reflection, scaffolding and feedback. This was demonstrated by their verbal reports of how the reflective prompts in eCREST influenced their data gathering strategies. Students commonly reported that eCREST helped them to structure their data gathering and remain focused on asking relevant questions to their diagnostic hypotheses. They found that it reminded them of the importance of having a broad initial differential diagnosis and continuing to consider and thoroughly investigate alternative hypotheses as new information becomes known. For students who adopted a 'Broad' strategy analytical thinking was perhaps not facilitated, as these students

reported that they selected questions and diagnoses in eCREST more randomly than those who used other strategies. More guidance may be needed throughout eCREST to support the development of a more 'Focused' data gathering strategy. However, repeated practice with online patient simulations could help students to develop a more 'Focused' strategy by receiving feedback and learning through trial and error. The benefit of using online patient simulations is that different approaches to gathering data and testing diagnostic hypotheses can be observed and educators can intervene when these strategies are inappropriate. Ideally, the online patient simulation could be used to detect weaknesses in students' strategies and provide constructive feedback to help them develop better strategies. This could perhaps be achieved by programming the simulations to calculate the proportion of relevant or essential information students gathered and their diagnoses, as I did for key feature one and two in my feasibility trial.

An interesting finding from my feasibility RCT was that the intervention group who had been exposed to eCREST had a tendency to take a 'Thorough' approach to gathering data i.e. they asked many questions, including many that were irrelevant. I hypothesised that these results could have been caused by the student's learning to be more cautious. eCREST presents complex general practice scenarios to students, in which patients present with common and vague symptoms. It also constantly prompts students to reconsider their diagnosis and provides feedback that emphasises how students should always consider serious conditions. These factors may have encouraged students to be more cautious and be 'Thorough' in their investigations. Taking a cautious

and 'Thorough' approach may be the most appropriate to take in these clinical scenarios. Therefore, prompting students to ask questions may be a good way of teaching reasoning skills, given that students have limited knowledge and experience and it is likely that through repeatedly practising asking questions they hone these skills over time, so that they ask fewer irrelevant questions. However, in real clinical practice it could also be impractical for them to ask many questions and have no benefit to patient outcomes (Berner & Graber, 2008; Newman-Toker, McDonald, & Meltzer, 2013). Indeed, studies have found that slowing down reasoning processes, or encouraging analytical thinking, does not necessarily lead to less missed diagnostic opportunities (Norman, 2009; Norman et al., 2014). Further consultation with GPs and medical educators is needed to fully understand the potential implications of this training on clinical practice. Future developers of online patient simulations and similar innovations need to consider what kind of clinical reasoning skills they are developing via their intervention and whether they are actually beneficial to students within real clinical practice conditions and for patient outcomes.

9.4 Implications

9.4.1 For policy

My findings have illustrated final year undergraduate medical students' understanding of how to manage and diagnose patients in primary care; particularly with patients who present with respiratory symptoms indicative of several potentially life-threatening conditions. My study focused on how students diagnosed patients but it implied there is a gap in students' skills in how to manage patients in primary care. This is perhaps not surprising given

that they have had very little experience of clinical management in primary care at this stage in their learning but this is a skill that they will need and develop in their next year of training as junior doctors in the NHS. Ensuring students are prepared and aware of how to manage patients like these in primary care is essential given the severity of the consequences if these symptoms are missed or not managed adequately. eCREST could help to fill this gap by allowing students to practise diagnosing and managing patients in this context in a safe environment. This could help to reduce the chance of future missed diagnostic opportunities, which may go some way to addressing policy concerns over patient safety and improving outcomes for patients (Illingworth, 2015; Yu A, 2016).

Further training using online patient simulations that focuses on how to identify and manage patients presenting with symptoms indicative of lung cancer would be particularly helpful for addressing policy concerns over the early detection of lung cancer in primary care (NHS, 2019). Evidence has shown that improvement in the early diagnosis and treatment of lung cancer has been much lower than other for other cancers; it is the leading cause of cancer death and is predominantly identified at an advanced stage through emergency presentation in secondary care (Bradley et al., 2018). Furthermore, the recent NHS Long Term Plan published in January 2019 outlines the NHS' plans to dramatically improve early diagnoses of cancer particularly through primary care routes by 2028 (NHS, 2019). If this goal is to be achieved, it is important to train future doctors as early as possible to recognise the signs of lung cancer

and other respiratory conditions in primary care and understand the most appropriate way to manage patients with those symptoms.

Current strategies focusing on targeting improvements in early diagnosis of lung cancer have partly focused on public awareness campaigns. For example, the 'Be Clear on Cancer' campaign to encourage early awareness of potential symptoms from patients, has shown to have increased the proportion of patients diagnosed with early stage lung cancer (Ironmonger et al., 2015; Peake, 2018). Other strategies to improve diagnosis, such as using risk assessment tools and decision support tools, have been shown to improve clinicians' decision-making and awareness by and increased referrals and diagnoses (Hamilton et al., 2013). Overall, a number of national cancer policies to improve early diagnosis in England over the past 12 years, such as the NHS cancer plan or the National Awareness and Early Diagnosis Initiative (NAEDI), have supported these efforts. However, there is evidence that these policies have had little impact on one-year survival of cancer patients and there are still substantial differences in cancer survival by socioeconomic status (Bradley et al., 2018; Exarchakou, Rachet, Belot, Maringe, & Coleman, 2018). Educational interventions targeted at future doctors, to improve their clinical reasoning skills, could provide an additional strategy to improving the early diagnosis and treatment of lung cancer and other life-threatening conditions. Online patient simulations like eCREST could help doctors to form the skills and awareness they need to recognise and treat conditions like lung cancer, from the beginning of their careers (IoM, 2015).

9.4.2 For practice

It is widely acknowledged that currently medical school curricula lack explicit ways of teaching clinical reasoning skills (Cleland et al., 2009; Higgs et al., 2008; Page et al., 2016). My findings showed that online patient simulations could be used to provide explicit teaching on clinical reasoning skills in medical schools. The results of my feasibility RCT suggested that medical schools perceived there to be value in teaching reasoning skills using platforms like eCREST but there remain issues on how best to implement this in a medical curricula and how best to evaluate them. In my feasibility trial, I found that medical students perceived an educational benefit in using eCREST as they felt it improved their reasoning. Moreover, my trial indicated that eCREST could potentially improve clinical reasoning by improving the way students gather information from a patient. It has provided proof of concept for a definitive full-scale trial to explore the effectiveness of online patient simulations on medical students' clinical reasoning skills.

Some researchers have warned against the replacement of face-to-face teaching with virtual methods of instruction, viewing uptake of new technologies to be a product of 'techno-romanticism'. Indeed, it should not be assumed that just because something is digital or online it is better than face-to-face teaching (Selwyn, 2013). However, there is a growing trend for the inclusion of more eLearning and simulation technology in medical education and in university education generally (Delgaty, 2015; Ellaway & Masters, 2008; Issenberg et al., 2005; Means, Toyama, Murphy, Bakia, & Jones, 2009). Online patient simulations may offer considerable benefits to faculty in freeing up time and

resources in the long-term, as they could reduce the need to organise interactions with simulated actor patients or small group discussions (PBL), content can be easily adapted and updated to suit the learner and they can be immediately delivered to students (Greenhalgh, 2001; Ruiz, Mintzer, & Leipzig, 2006). They also offer great benefits to students, as they can access content remotely, repeat courses more easily and receive a greater variety of simulated clinical experiences to learn from that might not be possible through other instructional methods (Greenhalgh, 2001; Ruiz et al., 2006). In medical education there are increasing demands for medical schools to provide additional course content to a greater number of students who may be geographically dispersed. Online patient simulations and other types of online learning may be the only way of feasibly delivering this teaching to a larger number of students (Greenhalgh, 2001). Moreover, the introduction of educational technologies does not necessarily mean that face-to-face teaching or other types of teaching will be replaced; the technology will likely be blended into the curriculum to support other teaching (Ellaway & Masters, 2008).

There are substantial barriers to introducing educational technologies into medical education. For example, there are likely to be significant short-term costs that can make the use of such technology implausible (Delgaty, 2015). This may become less of an issue if initiatives such as *electronic Virtual Patients* (eViP) become more commonly used, where a bank of virtual patients is available to be used for free by anyone (eViP, 2019). However, currently the cases available can lack key elements of simulation, such as interactivity, as students are not always required to gather information. The more pressing

barrier could be the organisational culture of medical schools and universities being resistant to change methods of teaching (Greenhalgh, 2001). My experience from conducting my feasibility trial in medical schools in the UK was that medical schools are fully equipped to use and interested in using online training resources, especially as most already use online custom learning environments, such as Moodle or Blackboard, to deliver current teaching and information to students. Furthermore, I found interest in continuing to use eCREST as a tool was high among the participating universities and has sparked interest from other medical schools in the UK. However, there do seem to be barriers to integrating innovations into a relevant module in an already saturated curriculum. Further research with medical faculties is needed to fully understand how educational technologies could be implemented into curricula and how to overcome the barriers to implementation.

9.4.3 For research

9.4.3.1 Description of the development and evaluation of eCREST

My systematic review showed that there was limited information on how to develop a theory based online patient simulation tool to improve clinical reasoning skills and the most appropriate way to evaluate such tools. Other researchers could benefit from my description of how eCREST was developed using theory, my development of an OPS logic model and my description of how eCREST was evaluated using a mixed-methods approach. This could guide future researchers to develop more theoretically and empirically informed interventions and evaluations that are more robust.

9.4.3.2 Full-scale RCT to evaluate eCREST

My research showed that it is feasible to conduct a definitive full-scale randomised trial using eCREST, which is adequately powered to assess the effectiveness of eCREST on clinical reasoning skills. There was good uptake and completion rates compared to other studies in which students volunteered to take part, as discussed in **Chapter 7**. Furthermore, medical students showed that they were highly satisfied with the tool; they felt it improved their clinical reasoning skills and felt it would be a valuable contribution to their curricula - demonstrating the intervention was acceptable to its target users. Based on the results from my feasibility RCT I estimate that a full-scale RCT, assuming similar uptake and completion to the feasibility RCT would need 256 students to detect an 8.3% change in the average proportion of essential information identified after one month, with $p < 0.05$ and 80% power. However, if a randomised crossover design were to be used a much smaller sample size would be required because there will be little within-participant variation (Sibbald & Roberts, 1998).

Future evaluations should consider several factors to ensure students are willing to use eCREST and take part in an evaluation, particularly if the tool is not integrated into the curriculum. For example, having champions who advocate and promote eCREST, from both faculty and students. If I were to conduct the study again I would attempt to engage with staff and students champions much earlier. I found it particularly difficult to identify a student representative to help champion eCREST. In retrospect, I should have spoken to some final year students before recruiting, to ascertain the appropriate

student contacts and identify where students are more likely to respond to advertisements to participate in research. Furthermore, I could have recruited my own student champion who may have helped me to advertise the study more directly to students. Indeed, this strategy proved very successful in my Think Aloud and interview study, where I used PALs medical students to help recruit students. Time of year when the trial takes place should also be considered. My feasibility RCT showed that uptake was greater when students received eCREST before their examinations; students also commented that they found it useful for revision purposes. Type of curriculum may also affect student engagement, as students in the PBL curriculum seemed to be less engaged than those in system-based curricula. Further research could look to recruit more medical schools that follow different approaches to teaching clinical reasoning to explore how online patient simulations can be integrated into different curricula. Indeed, more understanding of how best to integrate eCREST into medical school curricula in general is needed to determine whether it could replace or complement other instructional methods of teaching.

9.4.3.3 Further development of eCREST

My research indicated several ways in which eCREST could be developed, such as the inclusion of more patient cases, more explicit feedback on performance and more user input. eCREST is not seen as a static tool but one that is amenable to development and should respond to feedback from experts and users. With this in mind, my supervisors and I are intending to conduct some further research to consider how eCREST may be developed further. This will involve working with medical school faculty, GP registrars,

public health experts and students, to get their views on how best to implement eCREST into medical education curricula. It will help to determine what other patient cases would be particularly useful for medical students and how feedback should be presented to students and staff. My supervisors and I have also joined a project named EDUCATE (EDUCATE, 2018), which is a six-month training and development programme for educational technology groups. It aims to improve business and research skills to help educational technology projects develop further and will help the eCREST research team to explore how to implement eCREST into further medical schools in the UK and abroad.

9.4.3.4 *Further testing of measures for clinical reasoning skills*

My research indicated that the assessment of clinical reasoning skills is complex and there is unlikely to be a generic measure of clinical reasoning skills that can capture this set of skills. However, there are some validated approaches and frameworks for assessing reasoning, such as the key features problems that can be used to provide valid ways of measuring reasoning. Further research is needed to establish how to design these problems more quickly and efficiently, so they can be used for research purposes. Research is also needed to establish the reliability of these measures. Given that clinical reasoning skills are content and context specific - does performance on one set of key features for one patient case relate to that of another? Increasing the number of patient cases and their key features would improve the reliability and validity of the measure but it is unclear how many cases would be needed to ensure reliability and validity.

Future research could look to the key features structure that I took to assess clinical reasoning skills, which was to use a further online patient simulation as an assessment and identify key features of reasoning. The key features I identified were: how much relevant information was gathered, how much essential information was identified and how many times students adapted their diagnosis. These key features could be relevant for many if not most patient scenarios and could be used to assess reasoning in further studies. Further use of key features would help to validate this approach to measuring reasoning, as my study found that my approach could have poor predictive validity but other types of validity and reliability were not tested, as it was not the focus of this PhD.

9.5 Overall conclusions

This thesis contributes to the current literature by providing a logic model based on theory and empirical evidence to demonstrate how online patient simulations can help students to learn clinical reasoning skills and by conducting a methodologically robust evaluation of an online patient simulation. I found that eCREST was feasible and acceptable to medical students and suggested online patient simulations can improve the way students gather data by guiding them through the process. Further robust evaluations that integrate online patient simulations into the curriculum are needed to determine the effectiveness of such interventions on clinical reasoning skills. Future evaluations need to ensure that clinical reasoning skills are measured using valid and reliable measures to build a more robust evidence base for online patient simulations. If online patient simulations were found to improve medical

students' clinical reasoning when tested in a full-scale RCT, in which they were embedded into the curriculum, this could help students to potentially become better future doctors by improving the timeliness and accuracy of their diagnoses, particularly in primary care.

References

Association of American Medical Colleges. (2007). *Effective Use of Educational Technology in Medical Education Colloquium on Educational Technology: Recommendations and Guidelines for Medical Educators*. Retrieved from Washington, DC:

Aboshady, O. A., Radwan, A. E., Eltaweel, A. R., Azzam, A., Aboelnaga, A. A., Hashem, H. A., . . . Hassouna, A. (2015). Perception and use of massive open online courses among medical students in a developing country: multicentre cross-sectional study. *BMJ Open, 5*(1). doi:10.1136/bmjopen-2014-006804

Adams, A., Vail, L., Buckingham, C. D., Kidd, J., Weich, S., & Roter, D. (2014). Investigating the influence of African American and African Caribbean race on primary care doctors' decision making about depression. *Social Science and Medicine, 116*, 161-168. doi:10.1016/j.socscimed.2014.07.004

Adams, E., Goyder, C., Heneghan, C., Brand, L., & Ajjawi, R. (2017). Clinical reasoning of junior doctors in emergency medicine: a grounded theory study. *Emergency Medicine Journal, 34*(2), 70-75. doi:10.1136/emermed-2015-205650

Aghili, O., Khamseh, M. E., Taghavinia, M., Malek, M., Emami, Z., Baradaran, H. R., & Mafinejad, M. K. (2012). Virtual patient simulation: Promotion of clinical reasoning abilities of medical students. *Knowledge Management and E-Learning, 4*(4), 518-527. doi:10.3109/0142159x.2010.519066

Akyol, Z., & Garrison, D. R. (2013). *Educational Communities of Inquiry: Theoretical Framework, Research and Practice*: IGI Global.

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology, 10*(1), 67.

Arocha, J. F., & Patel, V. L. (1995). Novice diagnostic reasoning in medicine: accounting for evidence. *The Journal of the learning sciences, 4*(4), 355-384.

Asarbakhsh, M., & Sandars, J. (2013). E-learning: the essential usability perspective. *Clin Teach, 10*(1), 47-50. doi:10.1111/j.1743-498X.2012.00627.x

Audétat, M.-C., Dory, V., Nendaz, M., Vanpee, D., Pestiaux, D., Junod Perron, N., & Charlin, B. (2012). What is so difficult about managing clinical reasoning difficulties? *Med Educ, 46*(2), 216-227. doi:10.1111/j.1365-2923.2011.04151.x

Audétat, M.-C., Laurin, S., Dory, V., Charlin, B., & Nendaz, M. R. (2017). Diagnosis and management of clinical reasoning difficulties: Part II. Clinical reasoning difficulties: Management and remediation strategies. *Medical teacher, 39*(8), 797-801.

Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning. *Educational Psychologist, 40*(4), 199-209. doi:10.1207/s15326985ep4004_2

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology, 52*(1), 1-26.

Barrows, H. S. (1994). *Practice-Based Learning: Problem-Based Learning Applied to Medical Education.*

Basu Roy, R., & McMahon, G. T. (2012). Video-based cases disrupt deep critical thinking in problem-based learning. *Med Educ, 46*(4), 426-435.

Bateman, J. (2013). *Virtual patient design in undergraduate education.* University of Warwick. Retrieved from http://wrap.warwick.ac.uk/62110/1/WRAP_THESIS_Bateman_2013.pdf Available from University of Warwick Publications service WRAP database.

Bawa, P. (2016). Retention in Online Courses:Exploring Issues and Solutions—A Literature Review. *SAGE Open, 6*(1), 2158244015621777. doi:10.1177/2158244015621777

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American journal of medicine, 121*(5), S2-S23.

Beyer, B. K. (1997). *Improving student thinking: A comprehensive approach*: Allyn and Bacon Boston, MA.

Bezemer, J., Cope, A., Korkiakangas, T., Kress, G., Murtagh, G., Weldon, S.-M., & Kneebone, R. (2017). Microanalysis of video from the operating room: an underused approach to patient safety research. *BMJ Quality & Safety, 26*(7), 583-587. doi:10.1136/bmjqs-2016-005816

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: a social semiotic perspective. *International Journal of Social Research Methodology, 14*(3), 191-206. doi:10.1080/13645579.2011.563616

Bloom, B. S. (1953). Thought-processes in lectures and discussions. *The Journal of General Education, 7*(3), 160-169.

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy. *Medical Decision Making, 35*(4), 539-557. doi:10.1177/0272989x14547740

Bordage, G., Brailovsky, C., Carretier, H., & Page, G. (1995). Content validation of key features on a national examination of clinical decision-making skills. *Academic Medicine, 70*(4), 276-281.

Bordage, G., Grant, J., & Marsden, P. (1990). Quantitative assessment of diagnostic ability. *Med Educ, 24*(5), 413-425.

Botezatu, M., Hult, H., Tessma, M. K., & Fors, U. (2010). Virtual patient simulation: Knowledge gain or knowledge loss? *Medical teacher, 32*(7), 562-568. doi:10.3109/01421590903514630

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Med Educ, 40*(3), 254-262. doi:doi:10.1111/j.1365-2929.2006.02394.x

Bradley, S. H., Kennedy, M. P. T., & Neal, R. D. (2018). Recognising Lung Cancer in Primary Care. *Advances in Therapy*. doi:10.1007/s12325-018-0843-5

Braun, L. T., Zwaan, L., Kiesewetter, J., Fischer, M. R., & Schmidmaier, R. (2017). Diagnostic errors by medical students: results of a prospective qualitative study. *BMC medical education, 17*(1), 191.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Burke, H., & Mancuso, L. (2012). Social cognitive theory, metacognition, and simulation learning in nursing education. *Journal of Nursing Education, 51*(10), 543.

Calderhead, J. (1981). Stimulated recall: a method for research on teaching. *British Journal of Educational Psychology, 51*(2), 211-217. doi:10.1111/j.2044-8279.1981.tb02474.x

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ: British Medical Journal, 321*(7262), 694.

Caracelli, V. J., & Greene, J. C. (1993). Data Analysis Strategies for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 15*(2), 195-207. doi:10.3102/01623737015002195

Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & van der Vleuten, C. (2000). The Script Concordance Test: A Tool to Assess the Reflective Clinician. *Teaching and Learning in Medicine, 12*(4), 189-195. doi:10.1207/S15328015TLM1204_5

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal, 12*(2).

Cheraghi-Sohi, S., Holland, F., Reeves, D., Campbell, S., Esmail, A., Morris, R., . . . Singh, H. (2018). The incidence of diagnostic errors in UK primary care and implications for health care, research, and medical education: a retrospective record analysis of missed diagnostic opportunities. *British Journal of General Practice, 68*(suppl 1), bjgp18X696857. doi:10.3399/bjgp18X696857

Cheraghi-Sohi, S., Singh, H., Reeves, D., Stocks, J., Rebecca, M., Esmail, A., . . . de Wet, C. (2015). Missed diagnostic opportunities and English general practice: a study to determine their incidence, confounding and contributing factors and potential impact on patients through retrospective review of electronic medical records. *Implementation Science, 10*(1), 1-9.

Cleland, J. A., Abe, K., & Rethans, J.-J. (2009). The use of simulated patients in medical education: AMEE Guide No 42. *Medical teacher, 31*(6), 477-486. doi:10.1080/01421590903002821

Cochrane. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0 ed.).

Collins, A., Brown, J., & Newman, S. (1988). Cognitive Apprenticeship. *Thinking, 8*(1), 2-10.

Consorti, F., Mancuso, R., Nocioni, M., & Piccolo, A. (2012). Efficacy of virtual patients in medical education: A meta-analysis of randomized studies. *Computers & Education, 59*(3), 1001-1008.

Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine, 80*(6), 541-548. doi:10.1097/00001888-200506000-00005

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Academic Medicine, 85*(10), 1589-1602.

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hamstra, S. J. (2011). Technology-enhanced simulation for health

professions education: a systematic review and meta-analysis. *Jama, 306*(9), 978-988.

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D., Erwin, P., & Montori, V. (2008). Internet-based learning in the health professions. *JAMA: Journal of the American Medical Association, 300*(10), 1181-1196.

Cook, D. A., & Triola, M. M. (2009). Virtual patients: a critical literature review and proposed next steps. *Med Educ, 43*(4), 303-311. doi:10.1111/j.1365-2923.2008.03286.x

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37*(1), 45-54.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ, 337*, a1655.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. *Handbook of mixed methods in social and behavioral research, 209*, 240.

Croskerry, P. (2002). Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine, 9*(11), 1184-1204.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775-780.

Croskerry, P. (2009a). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in health sciences education, 14*(1), 27-35.

Croskerry, P. (2009b). A Universal Model of Diagnostic Reasoning. *Academic Medicine, 84*(8), 1022-1028. doi:10.1097/ACM.0b013e3181ace703

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf, 22 Suppl 2*, ii65-ii72. doi:10.1136/bmjqs-2012-001713

Davidoff, F., Dixon-Woods, M., Leviton, L., & Michie, S. (2015). Demystifying theory and its use in improvement. *BMJ Quality & Safety, 24*(3), 228-238. doi:10.1136/bmjqs-2014-003627

Delgaty, L. (2015). Twelve tips for academic role and institutional change in distance learning. *Medical teacher, 37*(1), 41-46.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials, 7*(3), 177-188.

Devitt, P., & Palmer, E. (1998). Computers in medical education 1: Evaluation of a problem-orientated learning package. *Australian and New Zealand Journal of Surgery, 68*(4), 284-287. doi:10.1111/j.1445-2197.1998.tb02084.x

Dewey, J. (1997). *How we think*: Courier Corporation.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447-456.

Department of Health. (2000). *An organisation with a memory*. UK: The Stationery Office Limited.

Department of Health. (2015). *The Government's mandate to NHS England for 2016-17: April 2016 to March 2017*. UK Retrieved from https://www.england.nhs.uk/wp-

content/uploads/2015/12/05.PB_.17.12.15-Annex-A-Mandate-to-NHS-England.pdf.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment:Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest, 5*(3), 69-106. doi:10.1111/j.1529-1006.2004.00018.x

Durning, S. J., Costanzo, M. E., Beckman, T. J., Artino, A. R., Roy, M. J., van der Vleuten, C., . . . Schuwirth, L. (2016). Functional neuroimaging correlates of thinking flexibility and knowledge structure in memory: Exploring the relationships between clinical reasoning and diagnostic thinking. *Medical teacher, 38*(6), 570-577. doi:10.3109/0142159X.2015.1047755

EDUCATE. (2018). EDUCATE. Retrieved from https://www.ucl.ac.uk/ioe/departments-and-centres/centres/ucl-knowledge-lab/six-research-themes/designing-smart-technologies-teaching-and-l/educate

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629-634.

Ellaway, R., & Masters, K. (2008). AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. *Medical teacher, 30*(5), 455-473.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). Medical problem solving an analysis of clinical reasoning.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med, 15*(11), 988-994.

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), 363.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review, 87*(3), 215.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*: MIT press Cambridge, MA.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Med Educ, 39*(1), 98-106.

eViP. (2019). eViP. Retrieved from http://virtualpatients.eu/

Exarchakou, A., Rachet, B., Belot, A., Maringe, C., & Coleman, M. P. (2018). Impact of national cancer policies on cancer survival trends and socioeconomic inequalities in England, 1996-2013: population based study. *BMJ, 360.* doi:10.1136/bmj.k764

Eysenck, M. W., & Keane, M. T. (2010). *Cognitive Psychology: A Student's Handbook*: Psychology Press.

Farmer, E. A., & Page, G. (2005). A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ, 39*(12), 1188-1194.

Fisseni, G., Pentzek, M., & Abholz, H.-H. (2008). Responding to serious medical error in general practice—consequences for the GPs involved: analysis of 75 cases from Germany. *Family Practice, 25*(1), 9-13. doi:10.1093/fampra/cmm071

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.

Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *British Journal of Educational Technology, 46*(3), 455-471. doi:doi:10.1111/bjet.12268

Garrison, D. R. (1991). Critical thinking and adult education: A conceptual model for developing critical thinking in adult learners. *International Journal of Lifelong Education, 10*(4), 287-303.

Gormley, G. J., Collins, K., Boohan, M., Bickle, I. C., & Stevenson, M. (2009). Is there a place for e-learning in clinical skills? A survey of undergraduate medical students' experiences and attitudes. *Med Teach, 31*(1), e6-12. doi:10.1080/01421590802334317

Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety, 22*(Suppl 2), ii21-ii27. doi:10.1136/bmjqs-2012-001615

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of internal medicine, 165*(13), 1493.

Green, S., & Higgins, J. (2005). Cochrane handbook for systematic reviews of interventions. In: Version.

Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education. *BMJ, 322*(7277), 40-44.

Groen, G., & Patel, V. L. (1985). Medical problem-solving: some questionable assumptions. *Med Educ, 19*(2), 95-100.

Hamilton, W., Green, T., Martins, T., Elliott, K., Rubin, G., & Macleod, U. (2013). Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract, 63*(606), e30-36. doi:10.3399/bjgp13X660751

Hamilton, W., Peters, T. J., Round, A., & Sharp, D. (2005). What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax, 60*(12), 1059-1065. doi:10.1136/thx.2005.045880

Harder, B. (2013). Are MOOCs the future of medical education? *BMJ, 346*, f2666.

Harding, A., Rosenthal, J., Al-Seaidy, M., Gray, D. P., & McKinley, R. K. (2015). Provision of medical student teaching in UK general practices: a cross-sectional questionnaire study. *British Journal of General Practice, 65*(635), e409-e417. doi:10.3399/bjgp15X685321

Hennink, M., Hutter, I., & Bailey, A. (2010). *Qualitative research methods*: Sage.

Higgins, J. P., & Green, S. (2005). Cochrane handbook for systematic reviews of interventions. In: version.

Higgs, J., & Jones, M. A. (2000). *Clinical reasoning in the health professions / edited by Joy Higgs and Mark Jones* (2nd ed. ed.). Oxford: Oxford : Butterworth-Heinemann.

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical Reasoning in the Health Professions*: BH/Elsevier.

Hogan, H., Healey, F., Neale, G., Thomson, R., Vincent, C., & Black, N. (2012). Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual Saf*. doi:10.1136/bmjqs-2012-001159

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277-1288. doi:10.1177/1049732305276687

Illingworth, John. Health Foundation. (2015). *Continuous improvement of patient safety: The case for change in the NHS*. Retrieved from London, UK:

Institute of Medicine. (2000). *To Err Is Human: Building a Safer Health System* (978-0-309-26174-6). Retrieved from Washington, DC: h[ttp://www.nap.edu/catalog/9728/to-err-is-human-building-a-safer-health-system](http://www.nap.edu/catalog/9728/to-err-is-human-building-a-safer-health-system)

Institute of Medicine. (2015). *Improving Diagnosis in Health Care* (978-0-309-37769-0). Retrieved from Washington, DC: h[ttp://www.nap.edu/catalog/21794/improving-diagnosis-in-health-care](http://www.nap.edu/catalog/21794/improving-diagnosis-in-health-care)

Ironmonger, L., Ohuma, E., Ormiston-Smith, N., Gildea, C., Thomson, C., & Peake, M. (2015). An evaluation of the impact of large-scale interventions to raise public awareness of a lung cancer symptom. *British Journal of Cancer, 112*(1), 207.

Issenberg, S. B., Mcgaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical teacher, 27*(1), 10-28.

Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5*(3), 235-283.

Johnson, P. E., Hassebrock, F., Duran, A. S., & Moller, J. H. (1982). Multimethod study of clinical judgment. *Organizational behavior and human performance, 30*(2), 201-230.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher, 33*(7), 14-26.

Kahl, K., Alte, C., Sipos, V., Kordon, A., Hohagen, F., & Schweiger, U. (2010). A randomized study of iterative hypothesis testing in undergraduate psychiatric education. *Acta Psychiatrica Scandinavica, 122*(4), 334-338.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist, 58*(9), 697.

Kahneman, D., Frederick, S., Kahneman, D., & Frederick., S. (2002). *Representativeness Revisited: Attribute Substitution in Intuitive Judgment Heuristics and Biases*: Cambridge University Press.

Kalet, A. L., Coady, S. H., Hopkins, M. A., Hochberg, M. S., & Riles, T. S. (2007). Preliminary evaluation of the Web Initiative for Surgical Education (WISE-MD). *American Journal of Surgery, 194*(1), 89-93. doi:10.1016/j.amjsurg.2006.12.035

Kamin, C., O'Sullivan, P., Deterding, R., & Younger, M. (2003). A comparison of critical thinking in groups of third-year medical students in text, video, and virtual pbl case modalities. *Academic Medicine, 78*(2), 204-211.

Kassier, J. (1983). Teaching clinical medicine by iterative hypothesis testing. *The New England journal of medicine, 309*(15), 921-923.

Kleinert, R., Heiermann, N., Plum, P. S., Wahba, R., Chang, D. H., Maus, M., . . . Stippel, D. L. (2015). Web-based immersive virtual patient simulators: Positive effect on clinical reasoning in medical education. *Journal of Medical Internet Research, 17*(11). doi:10.2196/jmir.5035

Kolb, D. A. (1984). *Experiential learning : experience as the source of learning and development*. London: Prentice Hall.

Kontopantelis, E., & Reeves, D. (2017). METAAN: Stata module to perform fixed-or random-effects meta-analyses.

Kostopoulou, O., Delaney, B. C., & Munro, C. W. (2008). Diagnostic difficulty and error in primary care—a systematic review. *Family Practice, 25*(6), 400-413. doi:10.1093/fampra/cmn071

Kostopoulou, O., Russo, J. E., Keenan, G., Delaney, B. C., & Douiri, A. (2012). Information Distortion in Physicians' Diagnostic Judgments. *Medical Decision Making, 32*(6), 831-839. doi:10.1177/0272989x12447241

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), 381-391.

Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf*, bmjqs-2015-004417.

Lee, A., Joynt, G. M., Lee, A. K., Ho, A. M., Groves, M., Vlantis, A. C., . . . Aun, C. S. (2010). Using illness scripts to teach clinical reasoning skills to medical students. *Fam Med, 42*(4), 255-261.

Lehmann, R., Thiessen, C., Frick, B., Bosse, H. M., Nikendei, C., Hoffmann, G. F., . . . Huwendiek, S. (2015). Improving Pediatric Basic Life Support Performance Through Blended Learning With Web-Based Virtual Patients: Randomized Controlled Trial. *Journal of Medical Internet Research, 17*(7), e162.

Lyratzopoulos, G., Vedsted, P., & Singh, H. (2015). Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *British Journal of Cancer, 112*, S84-S91.

Mamede, S., & Schmidt, H. G. (2004). The structure of reflective practice in medicine. *Med Educ, 38*(12), 1302-1308.

Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ, 42*(5), 468-475. doi:10.1111/j.1365-2923.2008.03030.x

Mamede, S., Schmidt, H. G., & Rikers, R. (2007). Diagnostic errors and reflective practice in medicine. *J Eval Clin Pract, 13*(1), 138-145.

Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Splinter, T. A., & van Saase, J. L. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res, 74*(6), 586-592. doi:10.1007/s00426-010-0281-8

Mann, J. A., & Roland, D. (2017). What are the measures that can be used to assess performance during in situ Paediatric Emergency Medicine Simulation? *BMJ Simulation and Technology Enhanced Learning*. doi:10.1136/bmjstel-2016-000140

Mann, K. V. (2011). Theoretical perspectives in medical education: past experience and future possibilities. *Med Educ, 45*(1), 60-68.

Maran, N., & Glavin, R. (2003). Low-to high-fidelity simulation–a continuum of medical education? *Med Educ, 37*(s1), 22-28.

Mayer, R. E. (2009). *Multimedia Learning* (2 ed.). Cambridge: Cambridge University Press.

McCoy, L. (2014). *Virtual patient simulations for medical education: Increasing clinical reasoning skills through deliberate practice.* Arizona State University Retrieved from

h[ttps://search.proquest.com/docview/1536437189/previewPDF/9FB3114](https://search.proquest.com/docview/1536437189/previewPDF/9FB3114E46D34A6APQ/1?accountid=14511) ProQuest database.

McDonald, P., Jackson, B., Alberti, H., & Rosenthal, J. (2016). How can medical schools encourage students to choose general practice as a career? *British Journal of General Practice*. doi:10.3399/bjgp16X685297

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2006). Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ, 40*(8), 792-797.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Med Educ, 44*(1), 50-63.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica, 22*(3), 276-282.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.

MedBiquitous. (2017). MedBiquitous. Retrieved from h[ttps://www.medbiq.org/](https://www.medbiq.org/)

Merriam, S. B., Caffarella, R. S., & Baumgartner, L. (2012). *Learning in Adulthood*: Jossey Bass Ltd.

Miller, M. R., & Levy, M. L. (2015). Chronic obstructive pulmonary disease

missed diagnosis versus misdiagnosis. *BMJ: British Medical Journal, 351*.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Monteiro, S. M., & Norman, G. (2013). Diagnostic reasoning: where we've been, where we're going. *Teach Learn Med, 25 Suppl 1*, S26-32. doi:10.1080/10401334.2013.842911

MOOC List. (2016). MOOC List Retrieved from h[ttps://www.mooc-list.com/](https://www.mooc-list.com/)

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition* (0309070368). Retrieved from

Nendaz, M., Gut, A.-M., Louis-Simonet, M., Perrier, A., & Vu, N. V. (2011). Bringing explicit insight into cognitive psychology features during clinical reasoning seminars: a prospective, controlled study. *Education for Health, 24*(1), 496.

Newman-Toker, D. E., McDonald, K. M., & Meltzer, D. O. (2013). How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Qual Saf, 22*(Suppl 2), ii11-ii20. doi:10.1136/bmjqs-2012-001616

National Health Service (2019). *The NHS Long Term Plan* Retrieved from h[ttps://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan.pdf](https://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan.pdf)

National Instiute for Health Research. (2016). Feasibility and Pilot studies Retrieved from h[ttp://www.nihr.ac.uk/funding-and-support/documents/RfPB/FAQs/Feasibility_and_pilot_studies.pdf](http://www.nihr.ac.uk/funding-and-support/documents/RfPB/FAQs/Feasibility_and_pilot_studies.pdf)

Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Med Educ, 39*(4), 418-427. doi:10.1111/j.1365-2929.2005.02127.x

Norman, G. (2009). Dual processing and diagnostic errors. *Advances in health sciences education, 14*(1), 37-49.

Norman, G., Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., . . . Monteiro, S. (2014). The etiology of diagnostic errors: a controlled trial of system 1 versus system 2 reasoning. *Academic Medicine, 89*(2), 277-284.

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Med Educ, 44*(1), 94-100. doi:10.1111/j.1365-2923.2009.03507.x

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2016). The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Academic medicine: journal of the Association of American Medical Colleges*.

Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: a review of the evidence. *Academic Medicine, 67*(9), 557-565.

Norris, S. P. (1985). Synthesis of research on critical thinking. *Educational Leadership, 42*(8), 40-45.

O'dowd, E. L., McKeever, T. M., Baldwin, D. R., Anwar, S., Powell, H. A., Gibson, J. E., . . . Hubbard, R. B. (2015). What characteristics of primary care and patients are associated with early death in patients with lung cancer in the UK? *Thorax, 70*(2), 161-168.

Offredy, M. (2002). Decision-making in primary care: outcomes from a study using patient scenarios. *Journal of Advanced Nursing, 40*(5), 532-541.

Offredy, M., & Meerabeau, E. (2005). The use of 'think aloud'technique, information processing theory and schema theory to explain decision-making processes of general practitioners and nurse practitioners using patient scenarios. *Primary Health Care Research and Development, 6*(01), 46-59.

Page, Bordage, & Allen. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine, 70*(3), 194-201.

Page, G., Matthan, J., Silva, A., & McLaughlin, D. (2016). Mapping the delivery of 'Clinical Reasoning'in UK undergraduate medical curricula.

Peake, M. Public Health England. (2018). *National cancer registration and analysis service. Be clear on cancer: regional and national lung cancer awareness campaigns 2011 to 2014. Final evaluation results—executive headlines.* Retrieved from London: ncin.org.uk/view?rid=3632

Pennaforte, T., Moussa, A., Loye, N., Charlin, B., & Audetat, M.-C. (2016). Exploring a New Simulation Approach to Improve Clinical Reasoning Teaching and Assessment: Randomized Trial Protocol. *Jmir Research Protocols, 5*(1). doi:10.2196/resprot.4938

Phillips, R. L., Bartholomew, L. A., Dovey, S. M., Fryer, G., Miyoshi, T., & Green, L. (2004). Learning from malpractice claims about negligent, adverse events in primary care in the United States. *BMJ Qual Saf, 13*(2), 121-126.

Posel, N., Fleiszer, D., & Shore, B. M. (2009). 12 Tips: Guidelines for authoring virtual patient cases. *Medical teacher, 31*(8), 701-708.

Posel, N., McGee, J. B., & Fleiszer, D. M. (2015). Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Medical teacher, 37*(9), 813-818.

Poulton, T., Conradi, E., Kavia, S., Round, J., & Hilton, S. (2009). The replacement of 'paper' cases by interactive online virtual patients in problem-based learning. *Medical teacher, 31*(8), 752-758. doi:10.1080/01421590903141082

Raupach, T., Muenscher, C., Anders, S., Steinbach, R., Pukrop, T., Hege, I., & Tullius, M. (2009). Web-based collaborative training of clinical reasoning: a randomized trial. *Medical teacher, 31*(9), e431-437.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *Jama, 298*(9), 1002-1009.

Round, A. (1999). Teaching clinical reasoning–a preliminary controlled study. *Med Educ, 33*(7), 480-483.

Rubin, G., Berendsen, A., Crawford, S. M., Dommett, R., Earle, C., Emery, J., . . . Gupta, S. (2015). The expanding role of primary care in cancer control. *The Lancet Oncology, 16*(12), 1231-1272.

Ruiz, J. G., Mintzer, M. J., & Leipzig, R. M. (2006). The Impact of E-Learning in Medical Education. *Academic Medicine, 81*(3), 207-212.

Rutherford-Hemming, T. (2012). Simulation methodology in nursing education and adult learning theory. *Adult Learning, 23*(3), 129-137.

Sandars, J. (2009). The use of reflection in medical education: AMEE Guide No. 44. *Medical teacher, 31*(8), 685-695.

Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign, 4*(1), 5-18. doi:10.1080/15710880701875068

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making, 7*(1), 1.

Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Med Educ, 49*(10), 961-973.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126): Basic books.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instr Sci, 26*(1-2), 113-125.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine, 8*(1), 18.

Selwyn, N. (2013). *Distrusting educational technology: Critical questions for changing times*: Routledge.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive learning environments, 15*(1), 27-46.

Sheikh, A., Panesar, S. S., Larizgoitia, I., Bates, D. W., & Donaldson, L. J. (2013). Safer primary care for all: a global imperative. *The Lancet Global Health, 1*(4), e182-e183.

Shephard, E. A., Neal, R. D., Rose, P., Walter, F. M., Litt, E. J., & Hamilton, W. T. (2015). Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case–control study using

electronic records. *British Journal of General Practice, 65*(631), e106-e113. doi:10.3399/bjgp15X683545

Sheringham, Sequeira, Myles, Hamilton, McDonnell, Offman, . . . Raine. (2016). Variations in GPs' decisions to investigate suspected lung cancer: a factorial experiment using multimedia vignettes. *BMJ Qual Saf*.

Sibbald, B., & Roberts, C. (1998). Understanding controlled trials Crossover trials. *BMJ, 316*(7146), 1719-1720.

Silk, N. (2000). What went wrong in 1,000 negligence claims. *Health care risk report, 7*, 13-15.

Simmons, B. (2010). Clinical reasoning: concept analysis. *Journal of Advanced Nursing, 66*(5), 1151-1158. doi:10.1111/j.1365-2648.2010.05262.x

Singh, H. (2014). Helping health care organizations to define diagnostic errors as missed opportunities in diagnosis. *Joint Commission journal on quality and patient safety, 40*(3), 99-101.

Singh, H., Giardina, T. D., Meyer, A. N., Forjuoh, S. N., Reis, M. D., & Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings. *JAMA internal medicine, 173*(6), 418-425.

Singh, H., Meyer, A. N. D., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf*. doi:10.1136/bmjqs-2013-002627

Singh, H., Schiff, G. D., Graber, M. L., Onakpoya, I., & Thompson, M. J. (2017). The global burden of diagnostic errors in primary care. *BMJ Qual Saf, 26*(6), 484-494. doi:10.1136/bmjqs-2016-005401

StataCorp. (2017). Stata Statistical Software: Release 15. In. College Station, TX: StataCorp LLC.

Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *Stata Journal, 4*, 127-141.

Sullivan, G. M. (2011). Getting Off the "Gold Standard": Randomized Controlled Trials and Education Research. *J Grad Med Educ, 3*(3), 285-289. doi:10.4300/JGME-D-11-00147.1

Swain, J. (2018). *A Hybrid Approach to Thematic Analysis in Qualitative Research: Using a Practical Example*. In. Retrieved from http://methods.sagepub.com/case/hybrid-approach-thematic-analysis-qualitative-research-a-practical-example doi:10.4135/9781526435477

Taylor, D. C. M., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical teacher, 35*(11), e1561-e1572. doi:10.3109/0142159X.2013.828153

The Society to Improve Diagnosis in Medicine. (2019). Clincial Reasoning Toolkit. Retrieved from https://www.improvediagnosis.org/clinicalreasoning/

Topol, E. (2018). The Topol Review. Preparing the Healthcare Workforce to Deliver the Digital Future. *Health Education England, editor. London, Unitied Kingdom: NHS*, 1-48.

UCL (2016). Case of The Month. Retrieved from https://www.ucl.ac.uk/medicalschool/teaching-portal/virtual/cotm

van den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library

catalogue. *Behaviour & Information Technology, 22*(5), 339-351. doi:10.1080/0044929031000

Vaona, A., Rigon, G., Banzi, R., Kwag, K. H., Cereda, D., Pecoraro, V., . . . Bonovas, S. (2015). E-learning for health professionals. *The Cochrane Library*.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.

Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: preliminary retrospective record review. *BMJ : British Medical Journal, 32*2(7285), 517-519.

Vygotsky, L. S. (1997). *The collected works of LS Vygotsky: Problems of the theory and history of psychology* (Vol. 3): Springer Science & Business Media.

Walters, S., Maringe, C., Coleman, M. P., Peake, M. D., Butler, J., Young, N., . . . Kölbeck, K. (2013). Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007. *Thorax*, thoraxjnl-2012-202297.

Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology, 92*(1), 3-10.

Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. (December 8 2016). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

West, C. P., Huschka, M. M., Novotny, P. J., & et al. (2006). Association of perceived medical errors with resident distress and empathy: A prospective longitudinal study. *Jama, 296*(9), 1071-1078. doi:10.1001/jama.296.9.1071

World Health Organization. (2017). *Technical series on safer primary care* (9241511680). Retrieved from Geneva:

Wong, G., Greenhalgh, T., Westhorp, G., & Pawson, R. (2012). Realist methods in medical education research: what are they and what can they contribute? *Med Educ, 46*(1), 89-96.

Wu, B., Wang, M., Johnson, J. M., & Grotzer, T. A. (2014). Improving the learning of clinical reasoning through computer-based cognitive representation. *Med Educ Online, 19*, 25940.

Yardley, S., Teunissen, P. W., & Dornan, T. (2012). Experiential learning: AMEE Guide No. 63. *Medical teacher, 34*(2), e102-e115. doi:10.3109/0142159X.2012.650741

NIHR Imperial Patient Safety Translational Research Centre. (2016). *Patient Safety 2030*. Retrieved from London, UK:

Zimmerman, B. J. (1995). Self-regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist, 30*(4), 217-221. doi:10.1207/s15326985ep3004_8

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal, 45*(1), 166-183.

Zimmerman, B. J., Boekarts, M., Pintrich, P., & Zeidner, M. (2000). A social cognitive perspective. *Handbook of self-regulation, 13*(1), 695-716.

Appendices

Appendix 1. Search history for MEDLINE, EMBASE, PsycINFO

The search was modified for ERIC, CINAHL, Scopus and Web of Science.

1.	Students, Medical/
2.	Education, Medical, Undergraduate/
3.	((medic* adj3 student*) and undergraduate).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
4.	Computer-Assisted Instruction/ or computer assisted instruction.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
5.	Teaching/
6.	(online adj3 (learn* or educat* or teaching)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
7.	exp Simulation Training/
8.	simulat*.mp.
9.	virtual realit*.mp.
10.	(patient* adj3 (virtual or simul*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]

11.	(elearning or e-learning or electronic learning or technology enhanced learning).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
12.	(learn* adj3 electron*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
13.	1 or 2 or 3
14.	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
15.	Clinical Decision-Making/
16.	Decision Making/
17.	(making adj3 decision*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
18.	decisionmaking.mp.
19.	((clinical or diagnostic) adj3 reasoning).mp.
20.	clinical judg?ment.mp.
21.	Judgment/
22.	(critical thinking or reasoning).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
23.	15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24.	13 and 14 and 23
25.	limit 24 to yr="1990 -Current"

Appendix 2. Steering committee members

Prof Stephen Duffy, QMUL Wolfson Institute of Preventive Medicine (Professor of Cancer Screening)

Prof Willie Hamilton, University of Exeter Medical School (Professor of Primary Care Diagnostics)

Dr Angelos Kassianos, UCL DAHR (Research Associate)

Dr Olga Kostopoulou, Imperial College London Department of Surgery and Cancer (Reader in Medical Decision Making)

Ruth Plackett, UCL DAHR (PhD Student)

Prof Rosalind Raine, UCL DAHR (Professor of Health Care Evaluation, Head of Department)

Dr Jessica Sheringham, UCL DAHR (Senior Research Associate)

Ms Raffaella Tate, Lay member

Dr Sanjiv Ahluwalia, (Head of School for General Practice Education in London, HEE NCEL)

Dr Anjali Bajekal, Friern Bernet Medical Centre (GP)

Dr Caroline Pelletier, UCL IoE (Senior Lecturer in Education)

Appendix 3. Further eCREST details

Login details to access eCREST

Website: http://silverdistrict.uk/ecrest/

Username: ▮▮▮▮

Password: ▮▮▮▮▮▮▮▮

Features of the clinical cases

	Features	
Case 1	**Key patient info**	John Roberts, Caucasian, male, 82 years old, retired.
	Initial presenting symptom	Long duration cough despite having flu jab. Family have also encouraged him to see the GP.
	History findings	- Cough - Reduced exercise tolerance - Recent cold - Reduced appetite
	Examination findings	- Mild bilateral pitting oedema
	Final Diagnosis	1. Lung cancer 2. Upper respiratory tract infection with post-viral cough 3. Upper airway cough syndrome (post-nasal drip syndrome) 4. Heart failure 5. Gastro-oesophageal reflux disease 6. Interstitial lung disease
	Management	Chest x ray
	Follow-up	Refer as a 2 week wait to the respiratory team
Case 2	**Key patient info**	Arjun Patel, Asian, male, 64 years old, security guard. He has history of hypertension, high cholesterol and pre-diabetes.
	Initial presenting symptom	Feeling tired all the time. Tired when wakes up.
	History findings	- Tired all the time - Uses 2 pillows to sleep with - Recent flu - Breathless - Reduced exercise tolerance - Swelling of feet
	Examination findings	- JVP elevated - Looks pale - *Bilateral pitting oedema of ankles* - *Proteinuria* - *BP slightly raised*
	Final Diagnosis	1. Heart failure 2. Recovering post influenza 3. Interstitial lung disease 4. Chronic Kidney Disease 5. Asthma 6. Diabetes 7. Vitamin D deficiency
	Management	Chest x ray

Bloods

Electrocardiogram and Echocardiogram

	Follow-up	Review in General Practice

Case 3	**Key patient info**	Taru Gandhi, Asian, female, 58 years old, dinner lady. History of sharing living space with family from India.
	Initial presenting symptom	Chest pains described as annoying and preventing her from doing everyday things
	History findings	- Pain is pleuritic - Productive cough for the past 2 weeks - Feeling breathless - She is a smoker
	Examination findings	- Cervical lymphadenopathy - Some tenderness over costochondral joints right side - Crepitations at the right midzone - Temperature 37.9 Low grade fever
	Final Diagnosis	1. Lower respiratory tract infection 2. Costochondritis 3. Pulmonary TB 4. pulmonary embolism (PE) 5. COPD 6. lung cancer
	Management	Electrocardiogram and Echocardiogram Relevant cultures or serological tests Prescription of appropriate medication
	Follow-up	Review in General Practice

Case 4	**Key patient info**	Zoya Akintola, African, female, 51 years old, cleaner. Smoker - 44 pack years.
	Initial presenting symptom	Requests flu jab as prone to chest inflections in winter and getting more breathless
	History findings	- Continuous productive cough - Clear phlegm - Feels breathless when cleaning and walking up stairs - Recurrent chest infections in winter - Breathless - Wheezy
	Examination findings	- Mild bilateral ankle oedema - Mild tachypnoea - Reduced expansion - Hyper-resonance to percussion, - Early inspiratory creps
	Final Diagnosis	1. COPD 2. Left heart failure 3. Asthma 4. Lung cancer 5. Bronchiectasis 6. lung cancer 7. interstitial lung disease
	Management	Chest x ray PEFR, spirometry and reversibly testing Bloods
	Follow-up	Review in General Practice

Appendix 4. Acceptability questionnaire

Please let us know what you thought about eCREST by rating how strongly you disagree or agree with the statements below on a scale from 1 to 5, using the following scale:

(1) Strongly disagree

(2) Disagree

(3) Neither agree or disagree

(4) Agree

(5) Strongly disagree

1. It was easy to navigate through eCREST

2. The level of difficulty of the material was appropriate

3. eCREST should be used to supplement traditional teaching

4. eCREST helped me to learn clinical reasoning skills that I could apply to my clinical work.

5. I would use eCREST in the future without an incentive.

6. What was your main reason for using eCREST?

   A) to be a better doctor B) to help me prepare for clinical practice C) to receive a voucher

7. Overall, using eCREST enhanced my learning.

8. Any suggestions for ways to improve eCREST: (open text)

Appendix 5. Self-report clinical reasoning measure.

Diagnostic Thinking Inventory, Flexibility in Thinking Subscale

Please read the following instructions before starting the survey:

The following questions are about your diagnostic thinking. Each item contains two accompanying statements and a rating scale. The scale refers to a continuum between the two statements. Please put a cross in the box which best describes your position on the continuum.

Do not try to work out any underlying meaning to each item; there is no right or wrong answer. Simply respond as spontaneously as you can by indicating how you actually diagnose and not how you think you should (even for those with little clinical experience). You will often find that you actually do things associated with both statements for a given item; your selection will indicate which one you do most often. If you hesitate between the two statements, please decide which one reflects what you do most often. You may think that there are other alternatives beside the two statements given (and there can be more than two in many instances), please make a choice on the basis of the two statements provided. It will take you about 5 to 10 minutes to complete the inventory.

	1	2	3	4	5	6	
1. When considering each differential diagnosis on my list,							
I try to prioritise the diagnoses							I try to give the diagnoses equal weighting
2. In thinking of diagnostic possibilities,							
I think of diagnostic possibilities early on in the case							First I collect the clinical information and then I think about it
3. When I am taking a history from the patient,							
I often seem to get one idea stuck in my mind about what might be wrong							I usually find it easy to explore various possible diagnoses
4. Throughout the consultation,							
If I follow the patient's line of thought, I tend to lose my own thread							I can still keep my own ideas clear even if I follow the patient's line of thought
5. When it comes to making up my mind about a diagnosis,							
I do not mind postponing my diagnostic decisions about a case							I feel obliged to go for one diagnosis or another even if I am not very certain
6. When I cannot make sense of the patient's symptoms,							
I move on and gather new information to trigger new ideas							I ask the patient to define those symptoms more clearly
7. When I am taking a history from the patient,							
I cannot bring myself to dismiss any information as irrelevant							I am quite happy to dismiss some information as irrelevant
8. When I cannot make sense of the patient's symptoms and signs,							
I can readily see the information in new ways							I find it difficult to see the information in new ways
9. When I cannot make sense of the patient's symptoms and signs,							
I move on to get new information							I try to reinterpret the data before moving on
10. When I am taking a history, I find that,							
I can get new ideas just by going over the existing information in my mind							I need to have new information to make me have a new idea about the case
11. When a piece of information comes along and							

makes me think of a possible diagnosis,							
It often makes me go back to previous information to see if things fit together or not							It rarely makes me review the information that I gathered previously
12. In relation to the diagnosis I eventually make,							
I usually have very few doubts							I often feel too uncertain for my own comfort
13. In <u>considering</u> diagnostic possibilities,							
I compare and contrast the possible diagnoses							I consider each diagnosis separately on its own merits
14. In <u>making</u> a diagnostic decision,							
I decide by considering each possible diagnosis separately on its own merits							I decide by comparing and contrasting the various possible diagnoses
15. As the case unfolds,							
I do not find it useful to summarize as I go along							I periodically take stock of the data and my ideas
16. When I have got an idea about what might be wrong with the patient,							
I feel most comfortable if I can follow it up without being diverted							I feel happy to go off on another tack and come back to my original ideas later
17. When I am taking a history from the patient,							
I manage to test my ideas even if I let the patient control the interview							I am only successful if I can control the direction of the interview
18. When it comes to choosing the most likely diagnosis from my list of differential diagnoses							
I usually find it difficult to rule out any of my diagnoses							I usually find it easy to rule out most of my diagnoses completely
19. Once I have made up my mind about a patient,							
I am prepared to change my mind							I really do not like to change my mind
20. When I am taking a history from the patient,							
I usually ask all the questions that I think are necessary during in the consultation							Quite often I do not ask all the necessary questions in the time
21. When the patient uses imprecise or ambiguous expressions,							
I let them go on to maintain the flow of the interview							I make them clarify precisely what they mean before going on

Appendix 6. The list of essential and relevant questions expected to be asked for each case

	Essential questions	Relevant questions
Case 1	When did your cough start?	Does your cough worsen on exercise?
	Can you describe your cough?	Does your cough worsen on lying down?
	Are you coughing anything up?	Have you had a cold recently?
	Have you coughed up blood?	Have you started a new job or spending a prolonged time in a new place?
	Do you have a wheeze?	What do you think might be causing this?
	Have you noticed any shortness of breath?	Have you recently started any new medication?
	Have you had a fever?	
	Have you had night sweats?	
	Have you noticed a hoarse voice or voice change?	
	Do you have any chest pain?	
	Do you or does anyone in your household smoke?	
	Have you had any weight loss?	
	How is your appetite?	
	Have you noticed yourself to be more tired than usual?	
Case 2	Has anything changed in your life recently?	Are you sleeping ok?
	How long have you felt like this?	Has your wife noticed anything abnormal with your sleep?
	Have you been unwell recently?	Do you fall asleep during the day?
	Do you have any other symptoms?	Do you have a fever?
	Do you find anything particularly difficult at the moment?	Do you take medication?
	How far can you walk?	Have you started any new medications recently?
	Do you snore?	Do you drink alcohol?
	How many pillows do you sleep with?	Is there anything else you are worried about?
	Are you more thirsty or passing urine more often than normal?	Did you have an ideas about what is going on?
	Have you noticed any bleeding?	

	Do you have any chest pain?	
	Do you having any breathing problems?	
	Has your bowel habit changed?	
	Have you noticed a change in your mood at all?	
	Have you lost or gained any weight?	
	Has your appetite changed?	
	Do you smoke?	
Case 3	Does anything make your pain worse?	Does your position alter the pain?
	Does anything make the pain better?	Have you had any injury to your chest?
	Does the pain come and go?	What do you do for a living?
	How long has the pain been there?	Have you started any new medications recently?
	Where do you feel the pain?	Do you drink alcohol?
	Does the pain travel anywhere else?	Have you lost or gained any weight?
	I can see you also have a cough. Can you tell me more about it?	
	Have you noticed any other symptoms?	
	When did the cough start?	
	Do you produce any phlegm?	
	Have you ever coughed up any blood?	
	Have you had a fever?	
	Are you more breathless than normal?	
	Any recent immobility?	
	Do you have pets?	
	Do you smoke?	
	How many cigarettes do you smoke per day?	
	How long have you been a smoker?	
	Do you have any ideas about what is causing this?	
	How is this affecting you?	
Case 4	Can you tell me more about the chest infections and the breathlessness?	For how long have you noticed the breathlessness?
	What type of work do you do?	What exercises can you do?

Have you noticed any other symptoms?

Do you ever cough up blood?

Do you cough up any phlegm?

Do you experience any chest pain?

Have you ever had any wheezing?

Do you feel breathless at night?

Do you need extra pillows to sleep?

Have you noticed your ankles or leg swelling up, in particular in the evening?

Have you lost any weight?

Have you travelled recently?

Have you experienced any night sweats at all?

Do you take medication?

Have you started any new medications in the past 6 months?

Do you have any allergies?

Do you smoke?

Is there anything else you are concerned about?

Do you currently have a cough?

Have you experienced any dizziness or collapse?

Have you been unwell recently?

Has your appetite changed?

Have you been around people with an infection?

Have you been exposed to any other chemicals than regular cleaning products?

Where are you from originally?

How long did you live there?

Appendix 7. Respiratory medicine knowledge quiz

Please note the appearance of these 20 questions will be randomised so that each participant will receive all the questions but not in the same order. In the pre-intervention online quiz 12 questions will be given to participants. In the month follow-up online quiz 12 questions will be given to students again consisting of 8 new questions and 4 repeated questions.

Question 1:

A 37 year old woman attends the surgery complaining of a dry cough. She has been coughing daily for some time and was told by a friend it was a chronic cough. She asks what is the minimum duration of symptoms that would mean she had a chronic cough?

What is the single most appropriate figure to give?

- 3 weeks
- 6 weeks
- 8 weeks
- 5 weeks
- 12 weeks

Question 2:

A 66 year old man has had a cough. He was referred by his GP eight weeks ago with a three week history of dry cough and runny nose but he did not attend his initial appointment. In clinic today he says the cough has completely disappeared.

What is the single most likely diagnosis?

- Chronic obstructive airways disease
- Exposure to cigarette smoke
- Gastro-oesophageal reflux disease
- Lung cancer
- Upper respiratory tract infection

Question 3:

A 60 year old man has a 6 week history of a dry cough, weight loss and digital clubbing. He is a non-smoker.

What is the first condition you would investigate this patient for?

- Bronchiectasis
- Chronic obstructive airways disease
- Lung cancer
- Post-infectious cough
- Tuberculosis

Question 4

A 70 year old female presents to you in General Practice with a cough for 9 weeks and increasing tiredness and appetite loss. She previously smoked around 25 cigarettes a day but gave up 10 years ago.

What Investigation would be your first priority?

- Spirometry
- Sputum culture
- Serum natriuretic peptide (e.g. BNP)
- Chest X Ray
- Gastroscopy

Question 5:

A 66 year old patient presents to the GP with a worsening cough. She started to feel more unwell last night and decided to come to see you today.

Which of the following additional pieces of information would NOT be an indication for referral to secondary care?

- CRB-65 score is 1
- Haemoptysis and weight loss
- Blood pressure 125/80, Pulse 84, Respiratory Rate 22, saturations 91% on air

- Suspicion of inhaled foreign body

- History of chemotherapy in the last 2 weeks

Question 6:

A mother brings her 5 year old boy to see you in a GP practice. The boy has been complaining of a cough and some difficulty in breathing. You consider the diagnosis of asthma.

Which of the following pieces of information would make asthma more likely?

- An audible wheeze during a viral illness

- A sibling with psoriasis

- Symptoms worse at midday

- Symptoms worse with exercise

- Associated prominent dizziness, light-headedness and peripheral tingling

Question 7:

You are a junior doctor working in general practice. A 65 year old gentleman presents to you stating he never has any energy, and feels tired all day.

Which of the following additional pieces of information would be most likely to trigger an urgent referral to secondary care?

- HbA1c 69 (normal range below 42 mmol/mol)

- Haemoglobin 11 g/dl (normal range 13.5-18 g/dl), MCV 79 (normal range 82-100 fl)

- eGFR 51 ml/min/1.73m2

- TSH 30 mu/l (normal range 0.5-5.5 mu/l), Free T4 2 pmol/l (normal range 9-18 pmol/l)

- ALT 80 (normal range 3-40 iu/l)

Question 8:

A 35 year old female patient presents to her GP with tiredness. She states she has been feeling tired all the time for the previous 2 months. She reports no other symptoms. Her GP decides to order some initial blood tests.

Which of the following would NOT be included when ordering initial basic blood tests to investigate her tiredness?

- Full Blood Count
- IgA tissue Transglutaminase (for coeliac disease)
- Thyroid Stimulating Hormone (TSH)
- Carcinoembryonic Antigen (CEA)
- HbA1c

Question 9:

A 50 year old female patient presents to the GP feeling tired. She has been reading on the internet about common causes of tiredness.

Tiredness is NOT common presenting complaint of:

- Obstructive sleep apnoea
- Anaemia
- Pre-diabetes
- Thyroid disease
- Depression

Question 10:

A 67 year old man has recently been diagnosed with right sided heart failure.

Which clinical features would be most fitting with this diagnosis?

- Bi-basal crackles on chest auscultation
- Raised JVP and pitting oedema to mid-shins
- Transient chest pain on exertion
- Bi-basal crackles on chest auscultation and a raised JVP

-   Wheeze brought on by exposure to cats

Question 11:

You see a 40 year old man with chest pain, which is better when he leans forwards. What is the single most useful initial diagnostic investigation?

-   Arterial blood gas
-   Bloods including CRP, Troponin and FBC
-   Chest X-ray
-   ECG
-   Peak flow or spirometry

Question 12:

You see a middle aged lady with pleuritic chest pain. Which single examination finding would make it appropriate to do an emergency secondary care referral?

-   Bilateral resonant chest percussion
-   BP 170/100 in both arms
-   Temperature of 38ºC
-   Tenderness over the chest wall
-   Unilateral leg swelling and tenderness over the posterior calf

Question 13:

You see a 30 year old man who fell whilst ice-skating yesterday and landed on the left side of his chest. He has had left sided pleuritic chest pain since. There is bruising and tenderness over his left lower ribs, breath sounds are normal, oxygen saturation 99%, respiratory rate 15 breaths per minute and heart rate 72 beats per minute. Which is the single most appropriate management option?

-   Advise regular analgesia and deep breaths
-   Arrange blood tests including an FBC and a clotting screen
-   Arrange a chest-X-ray
-   Arrange an ECG
-   Refer to A&E for further assessment

Question 14:

A 25 year old woman presents with a history of recurrent episodes of chest pain. They are associated with a feeling of impending death, palpitations, rapid breathing and tingling in her hands and feet. What is the single most likely diagnosis?

- Anxiety related chest pain
- Cardiac arrhythmia
- Recurrent pulmonary embolism
- Stable angina
- Thyrotoxicosis

Question 15:

A 30 year old man presents with a one day history of pleuritic chest pain on the left side. His trachea is deviated to the right. He has reduced breath sounds and there is a hyper-resonant percussion note on the left side. What is the single most likely diagnosis?

- Empyema
- Pleural effusion
- Pneumonia
- Pneumothorax
- Pulmonary fibrosis

Question 16:

Which is the single most appropriate prescription for a patient presenting to the GP with an infective exacerbation of COPD?

- Amoxicillin 500mg bd for 5 days
- Amoxicillin 500mg tds for 7 days
- Doxycycline 100 mg on the first day then 200mg od for 5 days
- Doxycycline 200 mg on the first day then 100mg od for 5 days
- Erythromycin 500mg/1000 mg od for 5 days

Question 17:

Which of the following symptoms is not typical in COPD?

- Chronic productive cough

- Diurnal variation in symptoms

- Exertional breathlessness

- Regular 'winter bronchitis'

- Wheeze

Question 18:

A 22-year-old male has had intermittent wheezing and shortness of breath for the past three months. He is a smoker and has a history of eczema and hayfever. His chest examination is unremarkable. Spirometry is arranged and reported as normal. What is the single most appropriate management of his symptoms?

- Baseline FEV1 repeated following inhaled corticosteroids

- Chest xray

- Peak flow Diary

- Refer to secondary care

- Trial of Salbutamol inhaler and low dose corticosteroids

Question 19:

You are asked to interpret the post-bronchodilator spirometry results of a 56-year-old woman who has been complaining of progressive shortness-of-breath

FEV1/ FVC = 0.60

FEV1 60% of predicted

- COPD

- Lung fibrosis

- Neuromuscular disorder

- Poor technique - repeat spirometry

- Sarcoidosis

Question 20:

A 34-year old lady who is 16 weeks pregnant presents to A&E with an exacerbation of asthma.

She receives nebulised Salbutamol and you, as a junior doctor, review her before discharge.

She has a good inhaler technique and her peak flow prior to discharge is 370 l/min (predicted

440 l/min). Her regular medication is a salbutamol inhaler (100 mcg) as required. What is an

appropriate action?

- Add inhaled salmeterol 50mcg bd

- Add inhaled ipratropium bromide 500mcg qds

- Add inhaled beclomethasone 200 mcg bd

- Make no changes to the medication

- Suggest to start salbutamol 100mcg qds

Appendix 8. Further quotes from medical students' with suggestions of improvement to eCREST

Comment	n	Representative quotes
eCREST was appropriate and applicable to clinical practice	9	"Very well designed program. One of the few that actually felt helpful in improving my clinical practice." (Student A)
		"It allowed me to train my thoughts to think like a clinician and showed the way one should put together signs and symptoms." (Student B)
		"A fantastic resource - I wish I could think of a helpful suggestion, but it really was outstanding." (Student I)
		"Would do it without a voucher if it was part of the curriculum, I do think it's most attractive for final year students though." (Student C)
		"Really enjoyed it. Good revision!" (Student J)
		"I think it would be a welcome addition to existing teaching or as a replacement for current SA." (Student L)
Motivated to use eCREST again without a voucher	4	"I did it for a voucher....but it was actually so informative. Would do it without a voucher if it was part of the curriculum." (Student C)
		"I think it was a very useful resource and I would definitely use it again, perhaps without incentive." (Student K)
		"It seemed to work very well and be at an appropriate level. Although I was initially drawn in by the offer of a voucher, I really appreciated the content." (Student L)
		"Although I was motivated to use eCREST to receive a voucher, I believe that it has a place to exist in the normal MBBS curriculum." (Student B)
Wanted more user input	5	"Before giving a dropdown list of possible differentials, have an open-ended box for the user to write down his/her own differential list. The user can then compare his/her differential list to those in the dropdown list at the end of the exercise." (Student D)

"Something I think that could improve it would be if users had to think of their own questions to ask and come up with their own diagnoses, tests." (Student K)

"I would like a difficulty setting to make cases harder. For example, instead of having 10 respiratory diagnoses that can be chosen and ranked, have diagnoses completely open (where students have to decide on the particular system that is affected." (Student M)

Wanted more feedback	4	"The discussion of the various possible diagnose was good but would have been useful to have more feedback on choice of investigations and management as this is what I struggle with at this stage more than the diagnosis itself" (Student E)
		"End of case summary sheet or a video of a fluid consultation." (Student N)
		"Clearer feedback on incorrect/missed differentials, and on questions that should have been asked in the summary PDF." (Student O)
Wanted more/different patient cases	11	"Design a range of difficulties, some with ambiguous diagnosis and some straightforward." (Student F) "Develop scenarios to test acute/emergency care management." (Student G) "More interactive. Or case based" (Student P) "More cases, not just rest." (Student R)
More user friendly	14	"Better way of keeping track of how many questions you have asked, so you don't forget and then are suddenly told to move to next step. This would mean that students pay more attention to the questions they are asking and have left to ask." (Student H) "Make the videos, wherever they are used, MUCH shorter and more concise. If you're relying on students learning in their own time, they want to learn as efficiently as possible." (Student S) "Please make it more obvious that there are many videos to watch at the end of each case study as I missed the first two having only watched the first video. The way that the text is highlighted didn't make it obvious to me." (Student T)

Appendix 9. Instructions for participants and topic guide for Think Aloud and interview study

Instructions for participants

Before we begin, let me just check that you have read the information sheet and given your consent to take part today?

Can I also check that you have completed the online quiz prior to taking part?

Have you ever taken part in a Think Aloud study before?

During this session, you will be given one simulated patient case to complete. I would like you to think aloud constantly while doing the task. This means that I would like you to complete the task, and while you do so, try to say everything that goes through your mind. For example, you might want to explain why you are asking the patient a question or why you are selecting each diagnosis. Please remember this is not a test of your ability and try not to worry about what you're saying or doing in the case. I'm not looking to see how good you are at diagnosing patients or grading your performance. I only want to know what's going through your mind when you use eCREST. Don't try to plan what you're going to say, just try and pretend that you're doing this at home and try to forget I'm here. I won't be able to respond to any questions you ask me during the task but feel free to say these questions out loud. It's very important that you keep talking. If you're silent for a long period, I will prompt you to talk. Please feel free to be critical in your thoughts and to express when things don't make sense, as this will help us understand how eCREST works or doesn't work. If you get stuck at any point, you can stop the case at any time and we can discuss it. Remember you are free to leave the study at any point and your data can be erased at your request.

Thinking aloud usually feels a bit strange at first as it an unusual task. Don't worry about it most people find it a bit unusual at first but quickly get used to it. At the end of the task we will have an opportunity follow-up on any of the questions you may have had and I will ask you a few

questions about what you thought of the task. I will be audio recording you as you do the task and taking some notes but these will all be anonymised and no one will be able to identify you from any quotes used in the write up.

We will start off with a practice task to make sure you feel comfortable. Do you understand what I want you to do?

Practice task

I would like you to find the Virgin east coast website and find a train from London to Edinburgh that leaves after 6pm tomorrow evening. Go through the booking right up until it asks you to sign and stop. Please make any decisions around price and coach selection as you normally would whilst trying to say everything that goes through your mind.

Do you have any questions about the think aloud?

Real task

I would like you to click on [insert patient case name] patient case in eCREST and complete this case by following the instructions on the website, gathering information from the patient and diagnosing and managing the patient. Please try and think aloud constantly while you are doing this, e.g. explain why you are choosing to ask questions and why you choosing certain diagnoses?

Think Aloud prompts and semi-structured interview topic guide

Remain silent for Think Aloud task. Prompts if students become silent for more than 5 seconds:

That's great. Keep talking

Why did you choose that?

Don't forget to tell me what you're thinking?

That's great, remember to keep speaking clearly

End of Think Aloud Task

Begin Semi-structured interview

Well done you've finished the task and you did really well, I know it's quite an unusual experience to think aloud constantly. Thank you for taking part. So now I'm going to ask you some questions on what you think about eCREST. Just to remind you that we are interested in both positive and negative thoughts and please feel free to be honest about you found it and what you were thinking and feeling throughout. Remember everything you say will be completely anonymised.

1. How did you find task?

Motivation

2. Would you be interested in using eCREST again and if so why or why not?
3. What was interesting to you about the task?

Prompt: uncertainty, that it was online, the videos, choosing differentials throughout.

Self-efficacy

4. How confident were you that you could complete the case?

Prompt: did you believe that you were able to complete the case?

Emotions

5. How did you feel when completing the cases?

Prompt: frustrated, confused, uncertain.

6. Did the feelings of the patient influence your decisions if so why or why not?

Handling task difficulties and demands

7. How did you find working through the case alone, without a facilitator/teacher?

Prompt: I noticed you asked me whether… Do you think it was difficult to complete alone?

8. Did you feel the level of difficulty of the cases was appropriate?

Prompt: you mentioned you didn't know this, or this seemed to confuse you, could you explain why?

Features of eCREST

9. How did you find the feedback on the differential diagnoses and management plans?

Prompt: did it help you to reflect on what you did and think about why you made certain decisions?

10. How did you find reviewing your diagnoses after every 6 questions?

Prompt: did it help you to think of alternative diagnoses, did it help you to think about what to ask next, did it help you to think about what tests you needed to do, did it make you feel more comfortable with the idea of uncertainty, did it make you to stop to think about why you thought about that diagnosis?

Last question now…

11. How did you find reflecting on your performance at the end of the case?

Prompt: did it help you to think how your assumptions led you to make certain decisions? Did it help you to identify what you could improve on?

Finally, thank you again for taking part. Do you have any final questions for me?

Here is your voucher, please sign this sheet to acknowledge your receipt of the voucher.

Appendix 10. Coding framework for Think Aloud and interview study

Name	Description and example quote
Phase 1 - Generating initial hypotheses (H)	Generating initial hypotheses after patient has given presenting symptom and read electronic patient record.
Concern (H1)	
Justifies (H1a)	Justifies concern or rates concern they have for patient based on background and presenting complaint e.g. "So I am writing that I am quite concerned. I'm not putting the most concerned because I haven't yet heard about the nature of the chest pain and I haven't asked her about the cough yet."
Changes mind (H1b)	Shows change mind about concern e.g. "Um, she may warrant hospital admission for that. Um, there's a sort of chest pain also. Um, so she could … actually I would probably maybe change mine to the very con … Yeah, I don't know. Well she's talking to me and she's clearly, you know, she's clearly come to the GP practice. So, she hasn't gone to A&E or everything."
Diagnoses (H2)	Thinks about diagnoses based on background and presenting complaint. They may suggest diagnoses without explanation, justify why they think it might be that diagnosis and order their diagnoses.
Justifies (H2a)	Explains how the symptoms fit their ideas of diagnoses or don't fit e.g. "pulmonary tuberculosis because she's from an Asian background."
Likelihood (H2b)	Refers to the order of their diagnoses on their list (we ask them to do this) e.g. "It might be pericarditis. Or stable angina. I'll put stable angina next but again she hasn't really said how the chest pain feels."
Primary (H2c)	Discuss their top differential or most likely e.g. "upper respiratory tract infection with post viral cough. I mean I guess that's probably the most likely"
Influence of eCREST (H3)	
Prompts H3a	Uses eCREST's list of differential diagnoses to inform initial diagnoses e.g. "Oh no that's not what I meant to do. Last one I'll put, none of the rest of these seem very likely." Or they go through each diagnosis given by eCREST ruling in or out
Visual/non-verbal cues H3b	Refers to taking information from video or non-verbal cues e.g. "Um, so it's like, I've noticed that she's got a cough as well."
Knowledge (H4)	Includes when student refers to what they know about risk factors or what they don't know e.g. "I can't remember if he said how many weeks it's been going on, but pending on the duration of the

		cough which I've now learnt eight weeks is chronic, erm, he might need an urgent referral or an urgent chest x-ray at least."
	Aware (H4a)	Student is aware of what they need to know before they can rule something in or out or to help them progress e.g. "She has got a smoking history. Uh, we don't know how long she smoked for"
	Don't know (H4b)	Student expresses a lack of knowledge about a symptom or diagnosis or test e.g. "Um, so, um, and her age as well, um, I'm not sure how that's also increased risk factor."
	Wrong (H4c)	Student repots incorrect knowledge, has misconception of risk factor, symptom, disease.
Next steps (H5)		Explains what information they might gather/need, explains what they would rule out/explains what treatments may be appropriate and how they would manage patient. e.g. "TB, I guess unlikely because he looks Caucasian and unless he's travelled somewhere where it's highly prevalent or lives in a certain place, erm, that could make it more likely, but again I'd have to ask him more questions about that, but relatively low at the moment."
Reflection (H6)		Includes when student says they find something difficult or helpful and steps back from doing the task to express an opinion or thought about what they're doing. e.g. "It's quite hard to order these without actually knowing anything else about the patient apart from chest pain."
Summarise (H7)		The student locates background information and previously accepted conclusions and merely repeats back information without interpreting.
	Interprets (H7a)	Interprets test results and may designate information as irrelevant e.g. "I don't think that's going to be very relevant. Hypertension for quite a few years and emergency surgery, again I'm not sure that's relevant."
	Repeats/clinical structure (H7b)	Repeats back information without explicitly interpreting e.g. "Okay. 57. Okay. So three years ago hypertension, high cholesterol seven years ago. HRT eight years ago. [Whispering]. Okay. History of depression, post-natal depression." Implicit interpretation occurring
	Risk factors (H7c)	Identifies important risk factors e.g. "But she has my attention because she has a smoking history and she's got chest pain and she's coughing. So I definitely want to investigate her further to make sure there's nothing sinister going on."
Phase 2a - Gathering information (G)		Student explains what questions they need to ask and what examinations and bedside test results they need to prove/disprove rule in/rule out diagnoses. Students ask questions and request test results.
Examinations and bedside tests (G1)		Includes any examinations and tests students order and the reasons they provide for why they are gathering information e.g. "Check if she's lost any weight."

	Justifies (G1a)	Explains why choosing test to rule in or out diagnoses or why not choosing test (doesn't just say that's important - needs to give reason why) e.g. "I want to examine her respiratory system because she's presenting with what sounds like a respiratory complaint."
	Justifies why not (G1b)	Explains why not asking a question e.g. "Genitals, musculoskeletal chest exam. Um, I guess that's more for costochondritis."
	Learned behaviour (G1c)	Selecting test as common to do in clinical practice e.g. "I actually think, abdomen, she hasn't really mentioned any pain in her abdomen. I think it would be good to check it generally, but probably not like the most important thing."
Influence of eCREST (G2)		
	Questions (G2a)	Includes where students refer to the list of questions or exams, or identify there wasn't a question available, or discusses number of questions left out of 6 e.g. "So I'm just clicking through to see some different questions that I can ask. Okay so okay fine." and "Okay so this cough in the video sounds quite dry and constant but she says that she produces phlegm. I don't really know how to take that because I can't really ask her any more questions about it."
	Visual/non-verbal cues (G2b)	Refers to taking information from video or non-verbal cues e.g. "Okay. So, she's pointed to her right-hand side."
Interpretation (G3)		
	Significance of results (G3a)	Interpreting significance information from questions they have asked the patient or from, including how concerned they are e.g. "Okay so at the bedside what would I like to do? Check O2 sats. 96 per cent so that's normal." Or "I don't think we … I really elicited that much new information."
	Refining differential (G3b)	Interprets information but explicitly discusses it's relation to diagnoses e.g. "I feel like, now she's got a productive cough, um, I guess it could still be … I'm not sure about whether my top differential would still be a PE."
Knowledge (G4)		
	Aware (G4a)	Student is aware of what they need to know before they can rule something in or out or to help them progress e.g. "I'll check her peak flow because she might have some degree of airway obstruction."
	Don't know (G4b)	Includes student saying they know or don't something or show uncertainty e.g. "Erm, I guess peak flow is the only other sort of respiratory test, but I don't know what the normal is if he's high then, but I mean it's quite low for a man, but yeah it's quite hard to tell when you've not got a baseline."

	Wrong (G4c)	Student repots incorrect knowledge, has misconception of risk factor, symptom, disease e.g. "Okay. So, she's pointed to her right-hand side. Um, it's quite unusual for ACS presentation or angina to present with a cough and right sided pain (G3). Generally, it's left sided pain."
Questions (G5)		Refers to any questions the students ask the patient and the reasons they gave for asking those questions e.g. "Does the pain come and go?"
	Justifies (G5a)	Explains why they are asking the question to rule in or rule out diagnoses e.g. "So I'm looking for if there's an infective cause. I'm going to check if she's had any recent immobility."
	Justifies why not (G5b)	Explains why not asking a question e.g. "I could ask her about whether the pain travels anywhere. I don't know if that would really help me."
	Implicit line of questioning (G5c)	Suggests questions which have specific logic i.e. being asked to rule ddx e.g. "She's … I'd like to ask her about weight loss."
Reflection (G6)		Includes reflecting on how they are gathering information. They might think about strategy and what questions are best to ask e.g. "Okay, erm … I feel like I'm being really unsystematic."
Summarise (G7)		Student repeats back information that they found out without explicitly evaluating it but implied student is structuring their thoughts e.g. "So about two weeks ago she noticed some pain."
Phase 2b - Reviewing diagnoses (D)		Comments when the students are reviewing their differential diagnoses during data gathering phase.
Diagnoses (D1)		Refers to when students mention a potential diagnosis. Also when students weigh up and evaluate the information they have gathered from questions and examinations and how they have ordered their diagnoses e.g. "Um, for now I'm gonna take off costochondritis. Um, I'm gonna take off ACS. She might have a pneumothorax."
	Justifies (D1a)	Explains why their diagnoses are likely by referring to information they have gathered or why certain diagnoses are unlikely e.g. "Angina seems less likely because it doesn't seem to change on exertion. So I might remove that."
	Likelihood (D1b)	Refers to how likely their diagnoses are and orders their diagnoses e.g. "So given that I'm now going to put the lung cancer one step higher."
	Primary diagnoses (D1c)	Discusses their most likely or top differential e.g. "I think post viral cough still has to be top just because he's had a recent cold and that is just the most common cause in common things that are common."
	Intuition (D1d)	Reports how their feelings have influenced decision-making e.g. "I feel that's lower on the list at the moment."
	No justification (D1e)	Does not justify diagnostic decisions e.g. "I think I can review you. Um, okay. So, no, I want to say my top differential remains the same. Um, and now I think um, I … I'm gonna go along to do her physical examination."

	Changes diagnoses (D1f)	Refers to changing diagnoses e.g. "I'm gonna put lung cancer a bit lower. She hasn't had weight loss."
Influence of eCREST (D2)		
	Prompts D2a	Where students refer to the list provided by eCREST of diagnoses by reading list out or where student considers new diagnosis from list e.g. "Erm, let's see if I've missed any. Erm, okay I think I'll leave it like that for now."
	Visual/non-verbal cues D2a	Refers to taking information from video or non-verbal cues e.g. "Um, so it's like, I've noticed that she's got a cough as well."
Knowledge (D3)		Includes when student refers to what they don't know or know or say they aren't sure and show uncertainty e.g. "Erm, I think that's, GORD is less likely as well because it's not worse lying down. I know that doesn't rule it out, but it just makes me think it's less likely."
	Aware(D3a)	Student is aware of what they need to know before they can rule something in or out or to help them progress e.g. "Erm, I think that's, GORD is less likely as well because it's not worse lying down. I know that doesn't rule it out, but it just makes me think it's less likely."
	Don't know (D3b)	Includes student saying they know or don't something or show uncertainty e.g. psychogenic cough. I'm not really actually sure what that is.. "
	Wrong (D3c)	Student repots incorrect knowledge, has misconception of risk factor, symptom, disease e.g. "Okay. So, she's pointed to her right-hand side. Um, it's quite unusual for ACS presentation or angina to present with a cough and right sided pain. Generally, it's left sided pain."
Reflection (D4)		Includes students feeling they are missing info or reflecting on how their reasoning changed in their final diagnoses from their initial diagnoses.
Summarise (D5)		Student repeats back information that they found out without explicitly evaluating it but implied student is structuring their thoughts e.g. "and I will just write a couple of sentences about why I put this down. So, on her history, she … I have found out that she has left sided … sorry, right sided chest pain which is pleuritic in nature."
Phase 3 - Managing the patient (M)		Refers to any further tests students would like to order and follow up of patient.
Follow-up (M1)		Says follow-up option and may explain why e.g. "I think she can be managed in primary care but I'd like to follow up the chest x-ray in six weeks when hopefully some antibiotics have cleared up the infection to make sure there's no underlying pathology."
	Justifies (M1a)	Explains why they are asking the question to rule in or rule out diagnoses e.g. "So I'm looking for if there's an infective cause. I'm going to check if she's had any recent immobility."
	Justifies why not (M1b)	Explains why not asking a question e.g. "Um, she's not really desaturating, so I don't think she needs and ABG."

	Learned behaviour (M1c)	Selecting test as common to do in clinical practice e.g. "I'm going to check her cholesterol levels just because you know, she's hypercholesteraemic and obviously, not totally relevant to this presentation but generally if she was coming in to the GP you'd want to just review that regularly."
Further tests (M2)		Says what tests will order and may explain why e.g. "So I'd like to do some bloods. I'd like to check if she's anaemic."
	Justifies (M2a)	Explains why they are asking the question to rule in or rule out diagnoses e.g. "So I'm looking for if there's an infective cause. I'm going to check if she's had any recent immobility."
	Justifies why not (M2b)	Explains why not asking a question e.g. "Um, she's not really desaturating, so I don't think she needs and ABG."
	Learned behaviour (M2c)	Selecting test as common to do in clinical practice e.g. "I'm going to check her cholesterol levels just because you know, she's hypercholesteraemic and obviously, not totally relevant to this presentation but generally if she was coming in to the GP you'd want to just review that regularly."
Influence of eCREST (M3)		Student looks at list of options regarding further tests or follow-up e.g. "Erm … What else? …"
Knowledge (M4)		
	Aware (M4a)	Includes whether student is aware of available tests and what's available in primary care. Refers to knowing or saying I don't know. e.g "Okay, so chest x-ray, erm, to rule out lung cancer and because I know there are like strict guidelines on who needs one and he's had a chronic cough so I think that's one of the criteria."
	Don't know (M4b)	Includes student saying they know or don't something or show uncertainty e.g. "psychogenic cough. I'm not really actually sure what that is.. "
	Wrong (M4c)	Student repots incorrect knowledge, has misconception of risk factor, symptom, disease e.g. "Initially I thought that the most likely diagnosis was a PE but since learning that she has a low-grade fever and a slight tachycardia, I feel that there is infective pathology underlying her presentation."
Reflection (M5)		Student reflects on reasoning and thinks about how they are making decisions e.g. "Oh, it's when my clinical reasoning falls down."
Final diagnoses (M6)		Student selects final clinical impression final diagnoses e.g. "Okay, fine, erm, so I'm going to move on … It seems quite dramatic to have lung cancer second, but I just don't really think it's any of the others so I'll stick with it."

	Justifies (M6a)	Explains why their diagnoses are likely by referring to information they have gathered or why certain diagnoses are unlikely e.g. "Angina seems less likely because it doesn't seem to change on exertion. So I might remove that."
	Likelihood (M6b)	Refers to how likely their diagnoses are and orders their diagnoses e.g. "So given that I'm now going to put the lung cancer one step higher."
	Primary diagnoses (M6c)	Discusses their most likely or top differential e.g. "I think post viral cough still has to be top just because he's had a recent cold and that is just the most common cause in common things that are common."
	Intuition (M6d)	Reports how their feelings have influenced decision-making "I feel that's lower on the list at the moment."
	No justification (M6e)	Does not justify diagnostic decisions e.g. "I think I can review you. Um, okay. So, no, I want to say my top differential remains the same. Um, and now I think um, I … I'm gonna go along to do her physical examination"
	Changes diagnoses (M6f)	Refers to changing diagnoses e.g. "I'm gonna put lung cancer a bit lower. She hasn't had weight loss."
Phase 4 - Feedback and reflection (R)		Stage in eCREST where students get feedback on their performance in the case and reflect on performance and reasoning
Immediate reflections from feedback (R1)		Students reflect on the feedback from GP and PDF and may compare their results to. Includes if students struggle to interpret feedback or references to not looking at feedback.
	Similar (R1a)	Students' responses are similar to feedback e.g. "Just having a look to see. Okay so the clinician has also put costochondritis. Yeah. I think that's kind of what I was putting under muscular skeletal. Okay."
	Dissimilar (R1b)	Student answers didn't fit with GP views and may explain why they disagree e.g. "He probably put more emphasis on PE than I would have probably done before I did the case."
	Knowledge (R1c)	Identified from feedback that they learned something new or identified a gap in their knowledge e.g. "Okay and I think the others are quite self-explanatory. Erm … Oh, maybe, I just want to learn about this actually. So I just want to look at this….VIDEO: postnasal drip."
	Justifies decisions (R1d)	Justifies why made diagnoses/line of reasoning etc. e.g. "I thought, you know, she had a PE. Um, and then I found out about her pleuritic chest pain so, I was initially thinking, oh, this definitely could be a PE. Um, and she had a cough um. So, I was ruling out some of cardiac thoughts about this pain. Um, but then, I sort of started to move on to a more infective um, like diagnosis."
Influence of eCREST (R2)		In post think aloud interview refers to comments about eCREST influenced their reasoning can be positive or negative.

	Student view was negative (R2a)		Explains how eCREST may have hindered reasoning or learning according to students e.g. "I probably should have had the lower respiratory tract infection higher up at the beginning. Yeah. I think I was misled because the actor sounded like she had a dry cough."
	Student view was positive (R2b)		Explain how eCREST influenced reasoning or learning in a positive way e.g. "I guess that you're kind of, you're expected to commit to things quite early on and then adjust them. That's quite good instead of just sort of going in to questioning thinking that anything could be wrong. You already start to think about narrowing things down. So that's quite good."
	Authenticity (R2c)	Limitations (R2ci)	Refers to fact that was a simulation so may have limited questioning, or choices e.g. "Yeah, I think I kind of forgot about this being like a GP setting."
		Realistic (R2cii)	Feels simulation was realistic e.g. "Yeah, I think I definitely think so. I think the sort of … um, the patient and her comorbidities and her um, kind of social situation were quite um, like uh reflected uh that sort of subset of patients quite well, you know."
	History taking (R2d)		How eCREST influenced questioning e.g. "I like didn't just like … didn't notice things on the … on the dropdown list or I didn't notice questions cos I was just …INT: Mm, mm. RES: … flicking through(F2a) um, trying to … yeah, just trying to skim read everything. And then I think there were quite salient questions or points that maybe I missed out."
	Diagnoses (R2e)		How eCREST influenced diagnoses e.g. "I liked the way that it encourages you to think about which way … which diagnoses are more relevant and less relevant and why you think so a certain thing."
	Pausing (R2f)		Refers to how eCREST gives time to pause and reflect e.g. "And I think balancing those things can, obviously, in your mind its quite a lot, but it's nice to have the time to just think about it."
	Uncertainty (R2g)		Discusses how eCREST makes them feel uncertain, also includes discussions of confidence e.g. "like a list of dropdown boxes and you rule one thing out. And then you're suddenly thinking about it again and you feel like quite unsure of what route you're actually."
	Knowledge (R2h)		Influence of eCREST on knowledge e.g. "I think for me it just served as more of a stepping to stone to be like oh okay, I should and read a little more about this."
Reflection on clinical reasoning in real life or future (R3)			Thinks on their approach to making decisions. How they would normally gather info from a patient or how they think they should reason after using eCREST. Mostly commented on reflection page question regarding what they will take forward from eCREST? e.g. "I guess revisiting a mental list would be helpful in a similar way to what we did throughout the questions. However, I prefer to finish one line of questioning before going back to look at my differentials as otherwise I tend to lose my train of thought."
Reflection on how they reasoned in eCREST (R4)			Talks about how they reasoned and chose diagnoses in their patient case and may also talk about how the patients' concern influenced decisions. Often occurs in reflection page when asked how their approach changed e.g. "Because I think I did put pericarditis higher than PE but

		I'm not really sure what my thought process was there actually looking back. Because I definitely would have thought that PE was more likely than pericarditis."
Reflection on performance (R5)		Talks about what they did well and need to improve on most often in response to these questions on reflection page
	Improve (R5a)	Refers to anything they need to improve on e.g. "I did feel like I could have excluded more earlier on. Like I didn't really have five I thought were likely."
	Well (R5b)	Refers to when student discusses what they did well e.g. "I kept several possible diagnoses sort of open in my head throughout so that I was able to use the results of the examination to narrow them down."
Feasibility of eCREST (F)		Thoughts they have on what they liked and disliked and usability taken from interview after Think Aloud has ended. Also includes comments on the functionality of eCREST that occur while doing the task e.g. technical errors or when students are unsure how to use eCREST
Confused (F1)		Any confusion over how eCREST works and what to press e.g. "Right what do I do next?"
Criticism (F2)		Aspects of eCREST that students did not like
	Features (F2a)	Features include feedback, reflection, reviewing diagnosis etc. e.g. "The like six questions and having to review is too rigid. Like because I kind of like to finish going on like okay so how's the pain."
	Usefulness (F2b)	Refers to if it is useful e.g. "Yeah. Yeah I'd use it again. But equally I feel like, it depends. If you're getting enough experience with talking to patient's in a GP surgery or in A and E or something you get the same kind of experience with asking for, asking the patient questions and trying to illicit the diagnoses."
Errors (F3)		Technical errors e.g. "It's not letting me scroll down."
Praise (F4)		Aspects of eCREST that students liked
	Features (F4a)	Features include feedback, reflection page, reviewing diagnosis e.g. "I think it's always quite useful to reflect on what you've just done because it makes you sort of learn from it better."
	Usability (F4b)	Includes comments on difficulty, realistic, easy to use, can be used alone, used again e.g. "Yeah. I think it's realistic that patients don't always present in exactly the way that you're expecting."
	Usefulness (F4c)	Comments on how it would be useful e.g. "I think it will be good for people who aren't able to sort of see patients by themselves very much"