The Creation and Validation of an Immersive Virtual Reality Simulator to Teach and Assess Clinical Reasoning

Anoop Krishna Prasad
Department of Surgery and Cancer
Imperial College London
United Kingdom

Supervisors:
The Late Professor Nigel Standfield
Professor of Vascular Surgery and Surgical Education
Imperial College London

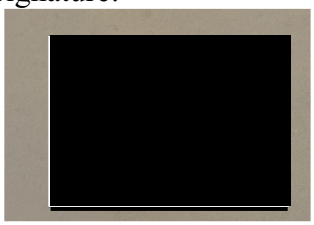
Dr. Stella Mavroveli
Senior Research Fellow
Imperial College London

A thesis presented for the degree of Doctor of Philosophy (PhD)

Declaration of originality

I hereby declare that the work presented in this thesis is original and I confirm that I have performed this research myself. Collaboration and important input from other individuals are declared in the respective chapters.

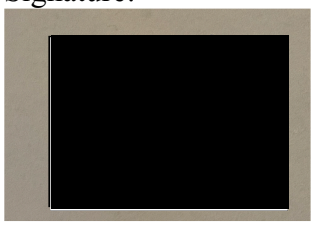
Signature:



Date: 20/11/2022

Copyright declaration

Signature:



Date: 20/11/2022

Acknowledgements

I must begin by thanking Professor Nigel Standfield for his support and guidance throughout my career. Whilst working with Prof as a core surgical trainee on educational material and techniques, he offered me the opportunity to pursue this PhD. Despite his busy schedule, Prof always made himself available, and was a constant source of support and advice, both on a personal and professional level. It was Prof's hope that my viva defence would occur face-to-face after the pandemic. Professor Standfield sadly passed away in March 2021 from a chronic illness, and the picture above is from our last socially distanced meeting. I will forever be grateful for his guidance, support and friendship.

I am extremely grateful to Dr. Stella Mavroveli, who picked up the reins as my PhD supervisor after Professor Standfield's passing. She helped to navigate the bureaucracy of changing supervisors and guided me tirelessly towards completing the submission. Stella's time, persistence and patient guidance have been invaluable in enabling me to get my thesis finalised, and I am grateful for all her support.

I would like to thank Dr. Simone Borsci, who's expertise in human factors testing was invaluable in planning our initial usability testing studies.

I owe a debt of gratitude to Matt Barrett and Florence Wilkinson, two medical students (now junior doctors) who helped greatly in the experimental work, both in the creation of cases, recording as patient actors, and in the assistance of the execution of the studies.

I would also like to thank Jernej Mirt and Arnaud Dressen, who both allowed me to use their technology platforms for the purpose of this thesis.

I am grateful to my parents for all their support, specifically my father, Raghavan, who was one of the patient actors, and did a fairly convincing job of it!

Finally, I would like to thank my wife Ahalya for all her love and support, for bringing me endless cups of coffee through many late nights in front of a laptop and for being a second pair of eyes for my thesis.

Abstract

Introduction

Emerging technologies such as virtual reality(VR) and 360-degree cameras have been presented as a technological solution that, through increased realism and presence, could lead to improved learning gains versus existing teaching modalities.

Aims

The main aim of the thesis is to create an immersive VR simulator using 360-degree camera technology to teach and assess clinical reasoning skills.

Methods

A multi-method approach across a series of studies were used to achieve the thesis aims. The first study aimed to develop a VR clinical scenario and scoring system using validated virtual patient guidelines, VR usability guidelines and scoring rubrics. A 360-degree camera and VR editing software were used for scenario construction. Three experimental studies were performed. Study 1: randomised controlled trial(RCT) comparing the simulator across two delivery modes - 2D screen(n=19) versus immersive head-set(n=20); assessing system usability and level of presence. Study 2: cohort study assessing the contrast validity of the simulator (medical students; n=16), foundation year doctors (n=21), and core surgical doctors (n=19). Study 3: RCT assessing learning gains (using Script Concordance test) with the simulator versus traditional methods(n=53).

Results

Four scenarios were created with scoring systems. The system was equally usable in both delivery forms(80.3 control vs 80 in treatment; $p>0.05$) with significantly higher levels of presence in the immersive group(0.81 vs 0.67; $p<0.05$). The simulator was able to differentiate between the most and least experienced group(57.3% vs 46.4%; $p<0.05$). Finally, the simulator learning gains were significantly higher compared to the traditional learning group(2.35 vs 1.64, $p<0.05$).

Discussion

This thesis demonstrates the feasibility of construction of a series of immersive VR clinical scenarios using 360-degree video. Taken together the results present a robust methodology for the creation and assessment of simulations using this technology and demonstrate the educational benefit for the teaching and assessment of clinical reasoning.

Thesis Outline

Recent challenges to medical education together with developments in multiple technologies have presented an opportunity for training. This thesis aims to create an immersive virtual reality ("VR") simulator using 360-degree video technology to teach and assess clinical reasoning skills.

Chapter 1 discusses the current healthcare training landscape, educational theories and clinical reasoning theory and research. The chapter also focusses on immersive VR, a history of its development, applications, and theoretical benefits for teaching and learning.

Chapter 2 introduces the development of a single immersive VR simulation, both proving its feasibility and allowing for further testing. Chapter 3 presents a study assessing the simulator using validated measurement instruments, assessing key factors including usability, cybersickness, face validity and presence. Importantly, it shows that the simulator had high usability scores, and that the immersive delivery mode resulted in higher levels of presence.

Chapter 4 describes the creation of further immersive VR clinical scenarios, and the development of an objective performance scoring system. Chapter 5 presents a randomised controlled trial (RCT) assessing the capability of the simulator and instituting a scoring system to differentiate between different levels of trainees (medical student, foundation year 2 trainee, core surgical trainee year 2). The simulator and scoring system were able to differentiate between all three groups, with the difference between the novice and experienced group reaching statistical significance.

Chapter 6 presents an RCT comparing the simulator against current training methods. It also details the literature and methodology for the creation of tools specifically designed to assessing clinical reasoning gains. The learning gains for the group taught using the simulator were significantly higher than those using traditional case-based discussion.

Finally, in Chapter 7, I present the discussion of thesis by objectives, the implications, limitations, and future directions to ensure this technology can be successfully implemented into current training.

Table of Contents

Table of Tables

Table of Figures

Word count – 46, 239

Chapter 1 - Introduction and Literature review

1.1 Medical education

The origins of modern-day medical education can be traced back to Flexner, who in his report to the Carnegie Foundation for the advancement of teaching, wrote a scathing report of healthcare education at the time (1). He wrote:

"*For twenty-five years past there has been an enormous over-production of uneducated and ill trained medical practitioners. This has been in absolute disregard of the public welfare and without any serious thought of the interests of the public.*"

He went on to state: "*Society forbids a company of physicians to pour out upon the community a horde of ill trained physicians.*"

At the same time, William Osler, was extolling the virtues of bedside teaching (2):

"*He who studies medicine without books sails an uncharted sea, but he who studies medicine without patients does not go to sea at all.*"

Changes since then led to a greater incorporation of science within medical education in undergraduate education (3) and a more formal structure to postgraduate training (4). By comparison, a physician in the 1930s entered independent practice 1-2 years after medical school versus a minimum of 8 years today (4).

1.1.1 Postgraduate medical education

1.1.1.1 Curriculum

There has been a concerted effort to move from a time/experience based approach towards a measurable outcomes based approach, due to the realisation that learning cannot be assumed to occur naturally during one's everyday work. Whilst this is still at an early stage (5), with uptake differing based on country and specialty, it has had a significant and positive impact (4). The key step is for the learning outcomes to be clearly identified and

communicated, which forms the basis of the curriculum, teaching methodologies used, assessment methodologies and training programme design (6). The curriculum also allows the trainee and trainer to gain a deeper understanding of training requirements, keep track of competencies gained, thereby allowing identification of areas in which the trainee is lacking.

1.1.1.2 Assessment

The definition of a clear curriculum allows for a more accurate assessment of competence and skill throughout training. This goes against earlier trends whereby the main form of postgraduate assessment was the membership or fellowship examinations by specific colleges, usually at the end of the training pathway.

The addition of trainee portfolios as an assessment tool encourages trainees to take charge of their own training. The AMEE guide (7) lists the advantages of this form of assessment as:

1. It counteracts the limitations of reductionist approach to assessment;
2. It facilitates the assessment of integrated and complex activities;
3. It supports an outcome-based approach to education; and
4. It actively involves the trainee and engages the trainee in the process of self-assessment.

1.1.1.3 Professional Teachers

The importance of education in training has been increasingly recognised in recent times, with both an increase in the training and assessment of teaching, as well as research into medical education. Courses such as “Train the Trainer” provide validation for the fact that those who educate require training in the skills required to teach (8). Trainers who used to be chosen based on seniority alone will have to demonstrate competence in teaching, with the GMC formally accrediting trainers (9). In addition, it has been increasingly realised that all

doctors, in one way or another act as educators. Efforts have therefore been made to incorporate teaching, and the assessment of teaching into postgraduate curricula.

1.1.1.4 Technology

In addition to the above, postgraduate education has, along with the rest of the world, begun to incorporate the rapid rise in various technologies from e-learning through to virtual reality technical skills-based simulators. The technologies relevant to this thesis will be discussed in more depth later in this chapter.

1.1.2 Challenges to medical education

Alongside changes in medical education, the healthcare system has changed, posing several new challenges to the delivery of effective medical education.

1.1.2.1 Reduction in training time

Since the introduction of the European Working Time Directive (**EWTD**) in August 2009, all postgraduate trainees in the UK have had a restriction in working hours to 48 hours per working week. The Temple Report of 2010 argued that given the right conditions, high-quality training could be delivered in 48 hours, though this would require a move to consultant-delivered care, with higher supervision levels, less out-of-hours work and more access to learning facilities (10). Whilst some specialties, such as anaesthesia have seen a minimal detrimental effect (11), craft-based specialties have shown significant reductions in the number of procedures completed versus pre-EWTD (12). It was shown that trainees in technical specialties more often work beyond the prescribed limit due to the need to maintain and advance their skillset (13).

1.1.2.2 Healthcare landscape

Perhaps a larger effect on training has been had by the gradual change in both the supply and demand of healthcare over the years. Data from the NHS between 2010 and 2000 shows a shrinking in the supply, or patient side of training. There has been a reduction in hospital stays from 7.8 days to 5.6(14)(3). There has also been an increase in elderly co-morbid patients who present reduced suitability for learning due to the complexity of their underlying conditions(3). Furthermore, there has been a push to see and treat more patients in an outpatient or community setting. Taken together, these changes have lead to a reduced number of training opportunities for junior doctors and medical students.

On the demand side, the government, and medical school have pushed to increase the number of medical student seats which has lead to a dilution of available training resources (15). In an effort to create mor learning opportunities, there has been a de-centralisation of clinical education to sites outside the academic environment, which presents additional challenges, such as access to more centralised learning resources (3). The overall effect of this reduction in supply and increase in demand is an environment that is increasingly challenging to the training of expert clinicians.

1.2 Educational theories

In the section that follows, we will be presenting education theories which are relevant to the topic of this thesis. It should also be noted that there is no single learning theory framework that incorporates how all adults learn in all different situations. Epistemology, the study of how we learn, is therefore not tightly bound to pedagogy, which encapsulates the theories surrounding the method and practice of teaching.

1.2.1 Constructivist and experiential learning

One of the most accepted theories of learning is the constructivist model (16), which, in its most basic form, supports the idea that new learning is based on building or connecting with what is already known. The fundamental underpinning of this, is that a learner creates a schema of the world around them based on previous knowledge, and then builds new knowledge by testing it against what is already known. Interestingly, this process has been described as identical in nature to the hypothetico-deductive reasoning/scientific method employed by clinicians during clinical reasoning (16).

In order to assist and supply a theoretical framework for the ideas and work presented in this thesis, I have briefly discussed below the key works in the field relevant to this thesis.

The underlying principle in the constructivist approach is that learning is an active process by which the learner adds information on to their pre-existing knowledge / experience. I have briefly outlined below the key proponents of this approach:

1.2.1.1 Lev Vygotsky

Vygotsky was a Soviet Psychologist dubbed the “Mozart of Psychology” (17) whose work has permeated through the fields of psychology, literary theory, neuroscience and psychiatry. Central to his approach was the belief that learning was social and cultural rather than just individual, so called, socio-cultural theories (18)(19). In this the social interaction is considered the primary focus, with the individual learning considered to be of secondary importance (19). The main contribution I will focus on is the idea of scaffolding and the zone of proximal development, which were both derived from his work on the way in which children learn. The zone of proximal development is described as the distance between the learner’s ability to complete a task independently and their ability to complete a task with

guidance. This is believed to be where learning occurs. It is helpful to consider these concepts as concentric circles. In the centre, there is what the student can do independently, the next layer out is what the student can do with support i.e. the zone of proximal development, and finally there is the outside layer, which is what the student cannot do even with support.

Scaffolding is the name given to the help and guidance provided to the learner by a more experienced individual. The support given to allow the learner to increase their learning can then be gradually removed, so as to allow the student to expand their individual competence in a task.

This principle has been applied in medical education and curricula by ensuring learning materials are presented at an appropriate stage in the learner's curriculum. For example, the initial stage of medical school involved introducing the fundamental sciences, following which learners are introduced to simulated patients and finally to practicing on real patients with pathology (20). This allows for students to ensure that they are gaining new knowledge, extending the zone of proximal development, but also ensuring that they are not pushed too far too soon so as to cause frustration in learning (20).

1.2.1.2 Donald Schon – reflection in action

Schon was a Professor at MIT in Urban Studies and Education whose work focussed on how professionals solve problems. It was a multi-faceted approach to the problem, looking at professions such as law, as well as medicine (21). Within medicine, he is most well-known for his work on reflective practice, for which he is considered a pioneer (22). The two main ideas relevant to this thesis are reflection-on-action and reflection-in-action. A novice would carry out reflection-on-action, that is, to reflect on an activity after it has occurred. Simply put, they would reflect on what happened, what was done, and how this could be improved next time

(22). An expert, Shon argued, can be defined by the ability for one to apply reflection-in-action, whereby one is able to generate a new understanding whilst the situation is occurring, using this to alter the outcome (22). These concepts have been adopted within post-graduate medical training whereby reflective practice is a core principle of learning and training. In addition, during teaching sessions such as simulations, as much time is given to the reflective aspect as the actual 'doing' of the task.

1.2.1.3 David Kolb

The experiential, or 'non-formal' learning theory of Kolb borrows ideas from Piaget and Vygotsky. This theory is rooted in constructivism as the 'concrete example' i.e. pre-existing knowledge underpins the cycle. Borrowing from Vygotsky, the social constructivist role of the teacher is important in enabling reflection and refining understanding (16). He defined learning as "*a process whereby knowledge is created through the transformation of experience*" (23). The cycle begins with a concrete experience or example, following which the learner tries to make sense of the experience through reflective observation (24). Abstract conceptualisation is the process by which the learner takes part in a "*transformation of that representation*" gathering and assimilating the various learning points and principles into existing knowledge (19). The final step in the cycle is where the new information is then tested in new experiences. This process does not exist without the help of a more experienced individual. Kolb commented on the fact that without appropriate support, learners may not be able to appreciate the nuances of the experience (19,23).

1.2.1.4 Miller

Miller created his pyramid, which was based on the original work by Bloom, who has a taxonomy which bears his name. It was created as Miller recognised that the assessment of medical students relied too heavily on what they knew, rather than how they would behave in the target environment (25). He therefore presented a four-tiered framework within which the assessment could occur. At the bottom of this pyramid was knowledge. Miller argued that whilst assessing knowledge was important, relying on this alone was an incomplete way of assessing one's clinical ability. The next level up is competence, i.e. knowing what to do, which can be assessed by clinical problem-solving exercises or extended MCQ's. His main argument was that achievement of the lower two levels, knowledge and competence, could not accurately predict the higher two levels of performance and action. Performance being described as a demonstration of learning e.g. in simulations or OSCEs, or an action describing how one performs in the real world e.g. workplace-based assessment.

The importance of this work cannot be understated, as it is seen to be responsible for the transition away from the Flexnerian model towards clinical performance-based assessments (26). It also highlights the lower two levels as being predominantly cognitive in nature, with the upper two levels being behavioural (25).

1.2.1.5 Ericsson

Whilst Ericsson's original work on deliberate practice has been cited heavily in the education literature, his original description was 'the individualised solitary practice in classical instrumental music as *directed by a qualified teacher*'(27). This concept has been widely studied in various domains, though using different definitions of the term, making drawing

conclusions across domains difficult(28). Despite this, the underlying work is worthy of discussion for this thesis, bearing in mind that the theory focusses on the acquisition of expertise, from a competent level of performance, rather than the stages of learning from novice to expert. Ericsson's original work was made as an argument against Galton's conceptualisation of expert performance being an innate ability or capacity, but rather something that was due an *'individuals prolonged effort to improve performance while negotiating motivations and external constraints'*(27). He argues that the key to expert level performance is to avoid stalling at a plateau when one reaches a level of automaticity with a given task. This requires not just extended experience, but a deliberate effort to improve(27). Deliberate practice, therefore, is a separate entity to just gaining experience, it is a structured activity explicitly aimed at improving performance, which requires effort, engagement and structured feedback to move the individual past the performance plateau that they are currently in. Ericsson himself states that medicine presents unique challenges in applying this expert-performance approach, however, this theory of expertise still has relevance in describing the importance of reflective learning and supervision throughout one's medical career(29).

1.2.2 Workplace learning

Experiential learning theories suggest authentic learning experiences become more important as the learners advance (19). Learning should therefore be embedded in practice, with the provision of authentic scenarios with relevant learning goals.

1.3 Assessment

The underlying theory on assessment used in medical simulation has been borrowed from early work in the psychological measurement literature. Assessment of tasks follows basic principles whereby the tests are deemed to be fair, reliable, valid and ultimately aligned to the task being taught or tested (30). Whilst this is an entire field in and of itself, I will summarise the main points relevant to this thesis below. It is of particular relevance as there is some disagreement in the use of these terms (31).

The American Psychological Association (32) proposed four ways of defining validity: content, construct, predictive and concurrent. Referred to as the four 'faces' of validity, they should be thought of not as individual and distinct entities, but as differing strategies to assess the same function, i.e. the meaning of a test score, with differing levels of significance based on the area in question. It is therefore now accepted that all efforts to understand the nature or validity of a test can be considered construct validation (30). Validation of a test can be used for two different purposes, firstly, it can be used to measure a specific characteristic in an individual, and secondly, it can be used to predict future performance of that individual. The first use is relevant to this thesis and will therefore be discussed further.

Face validity concerns the appearance of an assessment, i.e. does it look like it measures what it is meant to measure (33). In the medical education literature, this has been more simply described as 'the degree of resemblance between the system under study and the real activity (34). What is less often discussed, is its underlying importance. Increasing face validity, or appearing more realistic or similar to real-world applications can significantly improve the motivation of the subjects (33), which is of great importance when creating educational instruments. In establishing this, one must consider if it is more important to assess the novice or the expert with regards to this level of semblance. Again, the literature is not in complete

agreement (35), but the author feels that it should be assessed by the future potential users. This will be the approach taken in this thesis.

Content validity refers to the level to which the content of the assessment is pertinent and relevant to the subject being assessed (33). Again, in the medical education literature it has been used in differing ways, however, this author feels that it is justified to refer to it more as an expert's assessment of suitability of the tool (35–37). As such, this is more pertinent in the creation of the assessment tool than in some arbitrary validation after. In this thesis, the assessment (or clinical scenario) will be created with a team of expert surgeons who are aware of the aim of the tool.

In the psychological literature, construct validity is an overarching term that embraces several types of validity in describing how well the 'concept' underlying the test fits with what it is designed to do. In establishing construct validity, several studies must often be carried out, each assessing a different aspect of the assessment, forming in the end, an all-round approach covering many different aspects (33). Again, the medical literature has formed a consensus on one aspect of this definition, i.e. construct validity as the ability to differentiate between different levels of ability or experience (34). This author believes that the term 'construct' is an umbrella term that describes several of the other aspects of validity. For the description of the ability of an assessment to discriminate levels of experience, I will use the term contrast validity.

Concurrent and predictive validity are terms for which there is less disagreement. Concurrent validity simply assesses the extent to which performance on the assessment correlates with an established method of assessment. Finally, predictive validity assesses the extent to which performance on the assessment relates to performance with an established method of assessment in the future.

1.4 Clinical Reasoning

Despite the advances in medical education since Flexner, and the increased prominence given to patient safety, evidence has shown that the health-care system continues to be plagued by medical errors. The National Academy of Medicine in their 2015 reports Improving Diagnosis in Healthcare, stated that 'most people will experience at least one diagnostic error in their lifetime, sometimes with devastating consequences'(38). In the US, an estimated 40,000 deaths per year due to diagnostic error (39). Understandably, errors in diagnostic reasoning lead to greater harm than other is amongst the most important and therefore the field of clinical reasoning, the process by which a clinician gathers history, examines the patient and carries out appropriate tests, is gaining interest.

The concern of this thesis is to develop a tool to improve clinical reasoning. We will begin by defining it, discussing its importance and then present an historical review of the research thus far. Following this, strategies for developing clinical reasoning will be presented, along with advantages and disadvantages of each. Finally, the proposition for immersive Virtual Reality ("**VR**") will be discussed.

1.4.1 What is clinical reasoning?

Clinical reasoning is a complex process that describes the ability of a clinician to use available data from basic science, history, examination and investigations, in an environment of clinical uncertainty, to make decisions for patient management.

A more granular description of the process by Kassirer et al. (40) describes tasks such as:

1. Generating diagnostic hypotheses;
2. Gathering and assessing clinical data;

3. Deciding on the appropriateness of diagnostic tests;
4. Assessing test results;
5. Assembling a coherent working diagnosis; and
6. Weighing the value of therapeutic approaches.

1.4.2 Why is clinical reasoning important?

Clinical reasoning is an essential part of both medicine and surgery, with diagnostic errors accounting for more than 8% of adverse events in medicine and the majority of preventable errors in surgery occurring outside theatre (41)(42). Whilst diagnostic errors are commonly multi-factorial in origin, unsurprisingly, clinical reasoning issues accounted for 74% of the cases (43).

1.4.3 Research into the nature of clinical reasoning

Whilst most educationalists would agree that clinical reasoning is one of the most important skills of a clinician, despite 40 years of research, our understanding of this process is incomplete (44).

"It is unlikely that an area as broad as clinical reasoning will have a single construct that allows us to define an expert, rather expertise should be considered an amorphous entity whereby competent clinicians have an ability to adapt to the case presented by employing various tools available to them" Eva et al. (45).

Research into this field began in the 1970s with attempts to tease out this complex process using standardised patients and 'think-out-loud' protocols. Previous research had shown that students' problems-solving ability improved with education, it did not however show if this

was due to increased medical knowledge, increased exposure to similar cases, or in fact a bias due to the simulation method used in testing (46,47). Elstein et al. staged interviews between simulated patients (actors) and experts or medical students. During this time, they would verbalise their thought processes, using 'think aloud protocols'(48). Around the same time, another group led by Neufeld et al. conducted similar experiments with medical students of varying seniority. They gained insight into the clinical reasoning process through 'stimulated recall', whereby subjects would describe their thought process on viewing the recorded patient interaction (49). The research at the time demonstrated that doctors generated a limited number of hypotheses very early in the clinical encounter and that these had a great influence on subsequent data collection. Both studies gave rise to a model of clinical reasoning known as the 'hypothetico-deductive method'. This is where falsifiable hypotheses are generated, against which arguments are made using various findings. If the test outcome is contrary to the prediction of the hypothesis, the hypothesis is discarded. These studies showed that all subjects at all levels were performing the same process, the distinguishing factor between experts and novices being that experts made fewer and more accurate hypotheses (48,49). In addition, Elstein showed that success in one clinical case did not predict success in a subsequent case, termed 'content specificity'(48). He implied that content knowledge could not be readily separated from problem-solving ability, therefore, problem-solving ability was based on relevant knowledge to the specific problem. Subsequent studies also showed that the accuracy and promptness of hypothesis-generation played a significant role in the accuracy of diagnosis (49). Interestingly, work in other fields, such as chess, were unable to show gross differences in the mechanics of the thought process (50). De Groot was able to show that chess masters search through as many, or fewer (definitely not more)

possibilities, but were very good at selecting the right moves for consideration. In contrast, novices spent more time analysing worse moves (51).

Overall, this early work in medicine suggested that there was no generic problem-solving process associated with expertise. Experts appeared to generate hypotheses in the same way as novices (52), and the diagnostic accuracy and content specificity suggested the cause was down to underlying knowledge, rather than an expert reasoning process (44).

This work led to a shift in focus towards expert knowledge, rather than process. During the 1980s, work in related fields, specifically chess, suggested memory performance was the key to expertise. In short-term memory experiments, masters were able to reconstruct a chess game position after viewing it for only 5 seconds, and this ability was in stark contrast to players below master level. It was later shown that this was not due to the superior memory of the masters, as, when pieces were randomly placed on a board, there was no difference between masters and novices in their ability to reconstruct the board (50,51). The superior performance with meaningful games was based on their ability to 'chunk' what they saw based on the large set of game positions that they had learned through years of practice (50). Investigators attempted, with mixed results, to replicate these findings in medicine (52,53). Interestingly, use of clinical protocol showed no difference between the experts and novices, (54) however, incidental memory of lab values was a differentiator (53). An explanation for this was proposed by Schmidt, who suggested that experts have access to 'encapsulated' knowledge, only when it is needed (55). Because standard cases contain significant redundancy, this encapsulated knowledge does not emerge from simple recall tasks (55). Experts do however differentiate themselves when the tasks are more difficult, due to reduced time (55), or increased case complexity (54,55). Despite advancements in the

research, there existed discord between the theories of expert performance in other fields versus medicine. An explanation for the poor correlation between medical expertise and recall is that remembering extensive amounts of patient data is unlikely to be of benefit in most aspects of medicine (44). In contrast, knowing the positions of all the pieces on a chess board is critical to success during a game (56). It should be noted that this trend is not true in all fields of medicine. In renal medicine, for example, in which a large part of the task of the physician is to unravel and make meaning of various non-linear laboratory data, expertise has been correlated to recall (53,54).

The failure to explain expertise through recall methods and total amount of knowledge led to another change in direction (44). Perhaps the way in which experts organise or represent knowledge was the key. Experts in all fields including medicine, have a wider and deeper understanding of most topics as compared to novices. However, simply proving that they have more of a certain type of knowledge is not enough to prove that this is central to the development of expertise (44).

1.4.3.1 Illness Script Theory

In 1990, Schmidt et al. proposed a staged theory of clinical reasoning based on three assumptions. Firstly, on the road to becoming an expert, students progress through stages where they use different and evolving knowledge structures to solve problems. Secondly, these representations remain available for future use when called upon or 'activated'. Third, experts use knowledge structures labelled as 'illness scripts' when dealing with routine cases. These scripts emerge from, and are the result of, continuous exposure to patients.

Schmidt went on to describe four developmental stages, characterised by the emergence of four different knowledge structures (14):

Stage 1: The development of elaborated causal networks

Based on cognitive psychology models, the world is organised into cognitive models, which are mental representations of objects or events encountered. These 'propositional networks' can be imagined as a node which represents the concepts, and links which represent the relationship. Junior medical students presented with clinical cases would represent them in complex and elaborate ways based on the general underlying pathophysiology, with limited understanding of how the disease presents in the real world. This is understandable given that these causal networks have developed based on the material taught, which is mainly scientific and based on knowledge acquired from books (14).

Stage 2: The compilation of elaborated networks into abridged ones

With increasing exposure to patients, these declarative models are refined into high-level, simplified causal models that explain the constellation of signs and symptoms, which are given diagnostic labels (14). This process begins with the first patient exposure, and the knowledge use is continuously reorganised such that only knowledge pertinent to the case will be activated each time. This transition has been clearly shown when examining protocols of junior and senior medical students.

Stage 3: Emergence of illness scripts

The next stage involves the transition of refined causal networks into list-like structures, which Barrows et al defined as 'illness scripts'(57). Barrows described that physicians

reorganise knowledge of pathology, clinical presentation, signs and symptoms and conditions under which certain diseases, into scripts. The pathophysiological models used initially, are now compiled into simplified mental models, with sufficient underlying basic science to explain what is observed. This also explains the reason why although expert clinicians may have superior basic science knowledge, this is not mobilised in solving simple cases (44). Hobus et al. showed that, when asked to describe a typical patient with a disease, experts gave richer descriptions, specifically contextual factors that give rise to the disease (58). Clinicians develop varied illness scripts for the same disease based on their exposure, and interestingly, these bear only a superficial relationship to the prototypical cases in books (59).

Stage 4: Instance scripts / exemplars

The final stage in this process is the storage of memories of individual patients as memories, rather than a merged prototypical combination(14). In such cases, the expert is able to recall in great detail, incidental details, unrelated to the case, such as hair colour, over an extended period of months or years(14). This has been described as autobiographical memory and is described as a central feature of expertise in medicine. The illness scripts, the more generalised experience of a disease, including its pathophysiology co-exists in the mind of an expert with specific vivid recollections of previous patients and work together in a synergistic way (14).

The novice therefore moves through these stages, from causal networks, to illness scripts and finally instance scripts as they progress through their career. In this way they are able to carry out their role in diagnosis and treatment in a way that requires substantially less effort than having to work through a clinical problem actively or from first principles (14).

How do the illness scripts of experts differ from novices? Bordage has described the differences in how a patient problem is transcribed, with the use of semantic qualifiers (59). These describe bipolar representations of signs and symptoms, for example, acute versus chronic. Using this, he was able to classify four levels of expertise based on two dimensions, the semantic dimension (measured by the number of semantic qualifiers used) and the syntactic dimension (measure by the length of the summary) [28]:

1. Reduced discourses– limited SQ's and limited summary;
2. Dispersed – limited SQ's and extensive summary;
3. Elaborated – extensive use of SQ's with extensive summary; and
4. Compiled – extensive use of SQ's with limited summary.

These levels were related to diagnostic accuracy and can therefore be used as a proxy for expertise (59,60).

Summary

The four stages represent a developmental sequence in gaining expertise in clinical reasoning. Students start with an understanding of basic science, which on exposure to patients leads to causal networks which are further refined into illness scripts and exemplars. With increasing exposure over the years, these scripts are elaborated with specific instances which allow for rapid diagnosis and reduced mental effort (14). The expert can move between these levels based on the difficulty and familiarity with the case presented to them.

1.4.3.2 Dual process theory

Currently, the most accepted model of clinical reasoning is the dual-process framework, which has built further on Illness Script theory. The dual process theory postulates a fast, intuitive component (System 1) versus a slower more analytical component (System 2) (61).

As it is intuitive, System 1 is faster but also potentially more error-prone. As it characterised by the application of heuristics, it is the purview of the expert (62). Heuristics can be thought of as mental 'short-cuts' that reduce cognitive load. These simplifying pattern recognition heuristics used in System 1 are error-prone and a whole body of literature is dedicated to strategies to address this challenge (61). There is considerable evidence that clinicians use non-analytical processes when making diagnostic decisions (45)

System 2 is the more deliberate, analytical and conscious. Non-experts/novices employ System 2 thinking more often than experts, as they lack the experience required for System 1 thinking.

In reality, both forms are used and are not mutually exclusive (45). Eva described that "*the optimal form of clinical reasoning should be considered an additive model in which both analytic and non-analytic processes play a role*" (45).

Whilst the exact nature of the non-analytical mental models is unclear, it is assumed that it is either semantic networks (nodes of information units connected with other nodes in the network, the strength of the network and its nodes depends on the intensity of its use – resulting in schemas or illness scripts), prototypes (multiple encounters with related diseases leads to single prototype with common denominator) or instances (physician actually remembers individual patient encounters without abstraction).

1.4.4 How is clinical reasoning taught?

Without a clear understanding of the process of clinical reasoning, strategies to gain and teach clinical reasoning skill have posed a challenge. Research has shown clearly that clinical reasoning expertise cannot exist without appropriate content knowledge (45).

It has also been recognised that students have difficulty in applying foundational concepts to clinical problems, with only a 10-30% success rate in applying a concept to solve a new problem (63). This problem, termed 'transfer', relates to the ability to use knowledge acquired in one context to be applied in solving a dissimilar problem in another context. This issue gave rise to problem-based learning, which aimed to integrate basic science knowledge with clinical learning (63). Unfortunately, this approach, amongst others, has not shown any evidence of benefit versus the conventional curriculum, with some showing a worse outcome (64). Norman highlighted the importance of transfer - if only the principle is taught, the transfer to new problems is about 5% (65)(63). If the principle is illustrated with a single prototypical example, transfer may reach around 25%. If the principle is illustrated with multiple examples, transfer can reach 50% (66)(67)(63).

Different educational strategies have been shown to promote and develop clinical reasoning skills, which include:

a. Within the hospital – ward or outpatient setting: this allows for an authentic experience, although it brings with it difficulties associated with the current challenges to training, as well as raising the ethical issue of potentially putting patients in harms way;
b. Small group case-based discussions;

c. Simulation centre with mannequins and standardised patients: this is, however, resource-intensive in terms of both the equipment used and staff required. This is not currently a scalable resource and is inequitably distributed;
d. Virtual patients

Simulation has been proposed as a solution to this, more specifically, virtual patients, a modality within simulation, has been highlighted as the perfect tool to teach clinical reasoning (68). Virtual patients were initially developed in the 1970s due to the challenges of gaining clinical exposure due to a shift to outpatient care, higher student numbers and rising resistance for 'practicing' on patients (69). These are able to provide a series of patient exposures that allow for learners to build a database of examples.

The evidence has shown that teaching students a general process of clinical reasoning is likely to fail (70), and has shown that this is the case, in addition, the theory has shown that general reasoning strategies do not exist separate to the knowledge about a condition.

1.4.4.1 Virtual patients

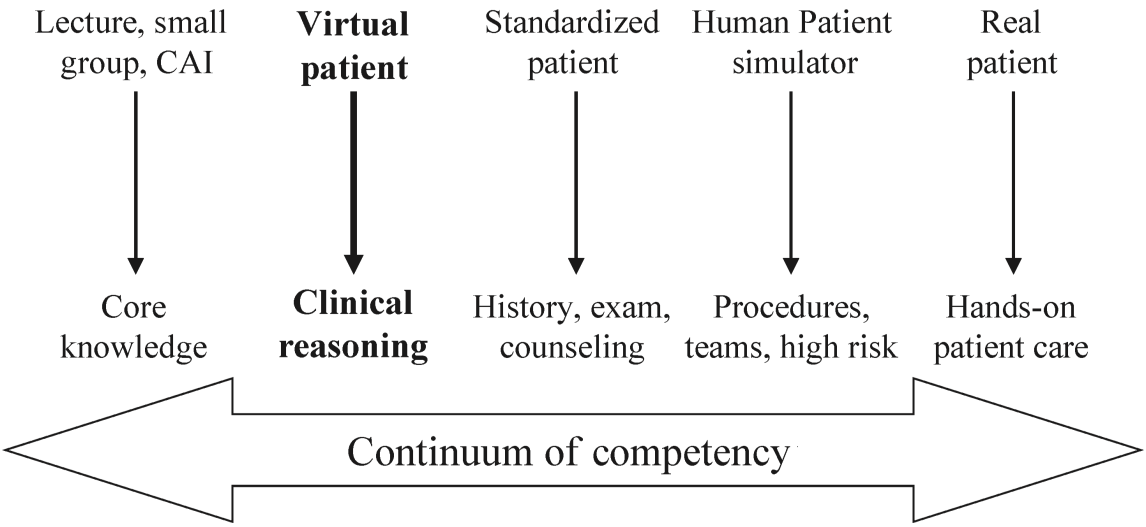


Figure 1: The continuum of competency (68)

While there is no universally agreed definition of a virtual patient (VP), the AAMC has defined it as "*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*".

Kenny et al. has described VPs as "*virtual interactive agents who are trained to simulate a patient's particular clinical presentation with a high degree of consistency and realism*" (71). The learner is able to proceed through a clinical encounter, by asking questions (usually through clicking on options), choosing relevant examination/tests, interpreting these and finally making a decision. It has been suggested that VPs are most well-suited to the teaching of clinical reasoning (68), and are less well-suited to developing soft skills for aspects such as history-taking, examination and communication skills (68).

The reason for this is that clinical reasoning ability is not a generic skill, as discussed earlier, but rather case-specific. In addition, if one accepts that expertise develops through the exposure to a large number of cases (with appropriate feedback and reflection), a large library of VPs is able to deliver this, and in a standardised fashion. The nature of VPs allows for the creation of similar cases, and this active learning with multiple examples has been shown to have an appreciable effect on a learner's ability to apply a previously learnt concept to a new problem. This is known as transfer (72).

In addition, VPs allow for deliberate practice, defined as (29):

1. Participants instructed to improve on a particular aspect of a task;
2. Immediate and detailed feedback; and
3. Opportunity to perform the task repeatedly

A fundamental aspect of this is specifically focussing on an aspect of performance (29).

1.4.4.1.1 Classification of virtual patients

Huwendiek et al. have attempted to typify VPs based on 19 factors synthesised into four categories (73):

1. general (title, description, language, identifier, provenance, and typical study time);
2. educational (educational level, educational modes, coverage, and objectives);
3. instructional design (path type, user modality, media use, narrative use, interactivity use, and feedback use); and
4. technical (originating system, format, integration, and dependence).

Bearman et al. classified VPs into two based on design, although both can be used in tandem (74):

1. Problem-solving design – the learner has to decide what information is relevant, i.e. the information is not cued. This primarily focuses on teaching clinical reasoning and diagnosing skills;
2. Narrative design – they suggest that this approach supports human interaction and increased value of the patient. This was found to be most useful when teaching communication skills.

Konocwicz et al. developed a classification system based on the technology used and the primary competency being taught/assessed (75).

Talbot et al. classified seven classes of VPs based on various factors:

1. Case presentation;
2. Interactive patient scenarios;
3. Virtual patient game;
4. High fidelity software simulations;
5. Human standardised patients;
6. High fidelity manikins;
7. Virtual standardised patient

VP forms vary from basic text-based scenarios to high-fidelity software simulations and virtual reality scenarios (75) and evidence has shown that interactive patient scenarios are can be used for clinical reasoning skills acquisition (68).

Despite the fact that the concept of VPs has been around for 40 years, they are present in only 24% of medical school curricula in the US and Canada (76). One of the main difficulties

associated with this has been cost, with the average cost of $10,000-50,000 per virtual patient.

Perhaps the best example of the current use of VPs is the computer-assisted learning in Paediatrics Program (CLIPP), which is used by more than 70 medical schools in the US (77) and has been demonstrated ability in filling in gaps in learners exposure to clinical cases (78).

In the next section, I will present advances in immersive virtual reality technology which have the potential for significantly advancing the teaching and assessment of clinical reasoning through virtual patients.

1.5 Immersive virtual reality

Immersive virtual reality has been heralded as the dawn of a new technological age with the current global virtual reality market size of 15.81 billion USD revenues in the billions of pounds (79). Venture capital companies have invested around $3.5 billion into 225 VR / augmented reality (AR) companies between 2014 and 2016 (80), with that figure steadily increasing. Recently, companies such as Facebook have set their sites on entirely immersive virtual world's with investment in the 'metaverse', which is touted as the next stage in the internet revolution(81).

By combining various technologies, including a head-mounted display, complex head-tracking systems with several degrees of freedom, spatial sound and ways to interact within this environment, VR technology now allows users to experience full immersion in a simulated environment.

The next section will discuss virtual reality, beginning with its definition, its use in healthcare, the history of its development, and end by discussing the theoretical argument for why it may allow the development of new teaching modalities that can improve clinical reasoning.

1.5.1 Virtual reality

Virtual reality is defined as a computer-generated simulation that replaces the real world using various media (82). There are four key elements in experiencing virtual reality (82):

1. Virtual world – this is the name of the content of a given medium, whereby a virtual world has been created. A modern example of this is Second Life [Figure2]. The way in which the virtual world is displayed defines if the experience is via virtual reality i.e. a virtual world can exist independently of its display through a virtual reality system. Virtual worlds, such as Second Life, can also be used to create massive multiple online role-playing games, whereby user interaction with peers within the world can lead to improved peer-to-peer learning, feelings of achievement and community (83)(84).
2. Immersion – the state of immersion can be generally described as an emotional or mental state, i.e. one could be described as being 'immersed' in a book or a movie. When describing immersion in VR, we are referring to a physical/sensory immersion rather than a mental one. This physical immersion itself does not have to include all senses.
3. Sensory feedback – the VR system provides feedback to the user based on their physical position which requires a high frame rate and reduced latency. This is achieved through tracking the user, usually through the head-mounted display or controllers and can involve varying degrees of freedom.
4. Interactive – the system should respond to the user's actions.

A concise definition of the term 'virtual reality' was provided by Sherman et al. who defined the concept as follow (82):

"*Virtual reality is a medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)*."

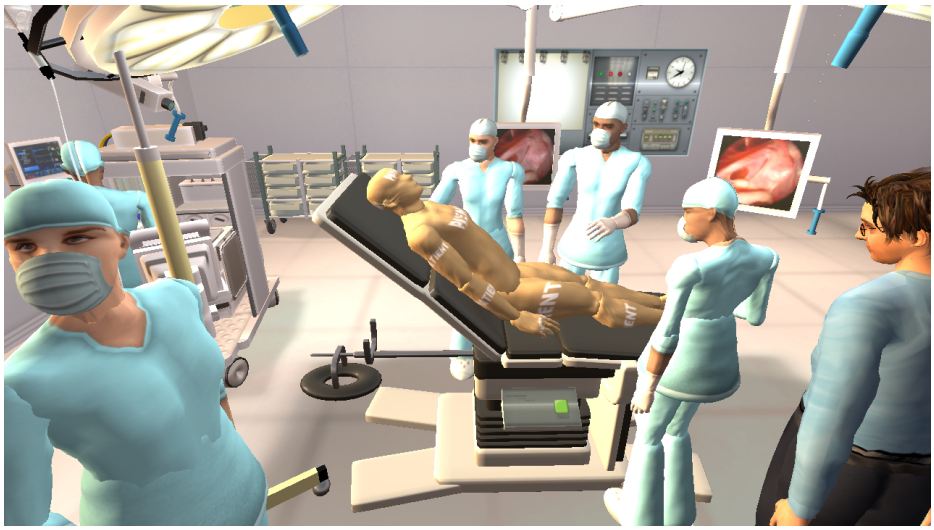


Figure 2: Second life is an example of a virtual world

The way in which this can be delivered can be through:

1. Head-mounted display –the participant is able to view images displayed on a screen, whilst the computer tracks movements of the user's head, allowing it to display appropriate images such that a user can look at the virtual world as they would the real world.
2. Cave automatic virtual environment – the virtual world is projected onto 3-6 walls of a room-sized cube. The user wears 3D glasses in order to view 3D graphics generated by the system.

Technology in this field has advanced rapidly in the last decade, with a significant reduction in price. Head-mounted displays are now available and affordable for the consumer market. These advances coupled with the reduction in price has led to an increasing number of applications within healthcare. To understand why there has been a recent increase in interest in this technology, it is important to highlight the history and previous limitations to its widespread adoption.

Figure 3: Global VR headset sales (Source – Financial Times) (83)

1.5.2 Why the interest in virtual reality in medical education?

There are two fundamental and interrelated principles that explain the interest in virtual reality as a potential tool to revolutionise the way medicine is taught. These are: (i) presence; and (ii) immersion, and I will describe each of these below (85).

1.5.2.1 Presence

Presence is described as 'the subjective experience of being conscious or the feeling of being spatially located in a particular place' (86,87). As highlighted by Rupp et al., the literature has suggested three factors that can create presence for the user (85):

1. The ability of the technology to physically surround the user in the experience (88,89);
2. The type of experience being simulated; and
3. The personality traits of the user.

The creation of presence is important as it allows for both increased recall and increased situational awareness within the simulation (90,91). Another reason is that due to the realism of the environment, users display behaviours that are more similar to their response in the real environment (92). This is of course linked to Bloom's taxonomy, whereby the assessment of users with this tool would cover behavioural aspects more than just cognitive. The previous point explains how immersion can also lead to improved levels of transfer in the real world. As the environment is more realistic, the user goes from requiring far-transfer, for a simple simulator, to near-transfer, for a more complex realistic simulator (85).

1.5.2.2 Immersion

As defined by Dede et al. '*Immersion is the impression that one is participating in a comprehensive, realistic experience*'(93). Simply put, Immersion is more a reflection of the technology used; it is an enabler of presence, which can provide higher fidelity. An example is of such use in surgery – it's action is through presence. By way of an example, if we compare a laparoscopic box trainer versus an immersive VR simulator with haptic feedback for the learning of triangulation skills, the immersive VR system would undeniably have a higher level of immersion, even though the simple box trainer would still be able to create the same level of presence for the task of triangulation. Therefore, even with a different level of immersion, the level of presence may be same, and therefore the learning may be similar.

Dede described that immersion is influenced by three main factors (93). Firstly, sensory immersion is the recreation of sensory experiences in the digital world, and this is improved

by the use of head-mounted displays, haptic technologies and stereoscopic sound. Secondly, actional immersion is the ability for the user to engage and initiate actions in the digital world. Thirdly, symbolic immersion is the use of everyday associations which can be used to build up context or provoke emotions, such as fear, even though in reality the user is completely safe. Whilst there is some disagreement about the exact definition, immersion is considered a more objective phenomenon related to the technology, i.e. with absolute sensorial immersion, the user's brain will be unable to tell the difference between the real world and the virtual world.

1.5.2.3 Why is immersion good?

There are two theoretical advantages to immersion that are relevant to this thesis. Firstly, virtual worlds allow for a change in the learner's perspective, which can be egocentric or exocentric. The egocentric frame of reference describes the view seen from within an object or space and has been linked with increased motivation and concrete learning (93). The exocentric frame of reference describes the view from the outside in, looking at an object, which has been shown to provide abstract insights as it gives the user perspective (93). An argument is made in the literature for combining these experiences into bicentric frames in order to capitalise on both strengths (93).

Secondly, immersion, in a similar mechanism to presence, may enhance the 'transfer' of knowledge to the real world. The transfer of knowledge can be described as near-transfer, where the learner has gained the ability to solve similar problems, and far-transfer, where the learner has gained the ability to solve problems in different contexts with different underlying semantics (93). A drawback of conventional teaching is the poor rate of far-transfer achieved. One potential significant advantage of immersive environments is their

ability to closely replicate the real world whilst the learner is practicing, which would potentially allow the user to apply this knowledge in the real world with only near-transfer, thereby increasing the effectiveness of the educational intervention. An often-cited example of this is a flight simulator (93).

1.5.2.4 Why is it better than video-based learning?

Video-based learning has gained popularity in teaching, with the development of several online courses. Despite the success, evidence has shown that because videos are less immersive, they can lead to less engagement, as compared to immersive virtual worlds which have been shown to increase engagement (94). There is however evidence that the learning gains from immersive technologies may be less, as the user is more present in the simulation and therefore distracted from the content being delivered (85).

1.5.2.5 The immersive learning landscape

Rapid advancements in technology have led to a host of technologies which sit along a continuum of immersion. Immersive virtual reality, at the highest end of the immersion spectrum, involves isolating the user from the real world by presenting a highly visual virtual world. This can be done using screens or a head-mounted display. Alternative reality platforms superimpose computer-generated images onto the real world, these can be interacted with and manipulated using a variety of inputs. This is achieved by overlaying images onto what the eye is seeing naturally, using glasses or projectors. A systematic review found that positive results had been shown in fields such as procedural training in surgery, the teaching of anatomy, however, studies have not shown positive outcomes for other applications such as general healthcare education, including clinical skills (95). As this field is rapidly evolving, it is likely

that the scope and breadth of its applicability will increase (95). A recent review by Turso-Finnich et al. into virtual reality head-mounted displays in medical education divided the current uses into: learning anatomy, training procedural skills, surgical procedures, communication skills and clinical decision-making (96). The review also notes that there are only four studies that have assessed the use of VR head-mounted displays in the clinical decision-making realm (96). The findings of these studies, along with their strengths and weaknesses is presented in Table 1.

Summary of studies					
	Study	Study description	Participants	Strengths	Weakn
1	Harrington et al. (97)	Feasability study - Computer-generated image (CGI) immersive VR trauma simulator (Advanced Trauma Life support scenario)	n = 30 19 - trainees 11 - course instructors	-Able to distinguish between differing levels of expertise based on performance score. -Use of introductory module	-No co modali -Perfor correla -No de for con -No ass level o -No jus used
2	Lowe et al. (98)	Cross-sectional observational study at a National Conference us a 360 degree video for a paediatric mass casualty incident	n = 207 63 <30 years old 85 - 31-40 years old 33 – 41-50 years old 24 >50 years old	-Large number of participants -Use of introductory module	-No co modali -Differ used w -Inclus non-ph occupa -Lowe particip -No de for con -No ass level o
3	Aussedat et al (99)	Pre-post test study of ENT residents using a CGI VR simulator	n = 22	-Statistically significant improvement in test scores following use of simulator. -Use of face validity and content validity instruments	-No co trainin -No de for con -No ass level o -No jus used

4	Zackoff et al. (100)	Prospective RCT comparing virtual reality head-mounted training in addition to standard training	n = 168 medical students Control n = 90 Intervention n = 78	-Prospective RCT -Large number of participants -Scoring rubric calculated using a validated method	-This v two gr group teachin than th explain -No de for con -No as level o

1.5.3 History of immersive virtual reality

The first patent for a head-based display was awarded in 1916 to Albert B. Bratt, for a head-mounted display designed as a periscope. In 1956, Morton Heilig developed the Sensorama, the first immersive, multi-sensory device, which was based on cinema theatres at that time (101). This 'experience theatre' concept lead to a prototype built in 1962, known as the Sensorama. Advanced for its time, it was able to display stereoscopic 3D images in wide-angle, with stereo sound and odour emitters, in addition to a motional chair (101). The first experience was a bicycle ride through Brooklyn, unfortunately, due to funding difficulties, development was halted a few years later (101). Whilst fully immersive, this was not interactive and not head-mounted.

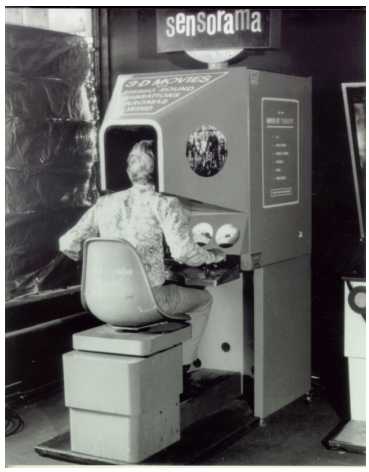


Figure 4: The Sensorama (84)

In 1961, engineers Comeau and Bryan created a head-mounted display that would remotely track and display viewing from a nearby video-camera (82). A major breakthrough occurred in 1965, when Ivan Sutherland, a student at MIT, introduced the concept of the 'Ultimate Display'. He described a display that would allow the user to interact with objects not bound

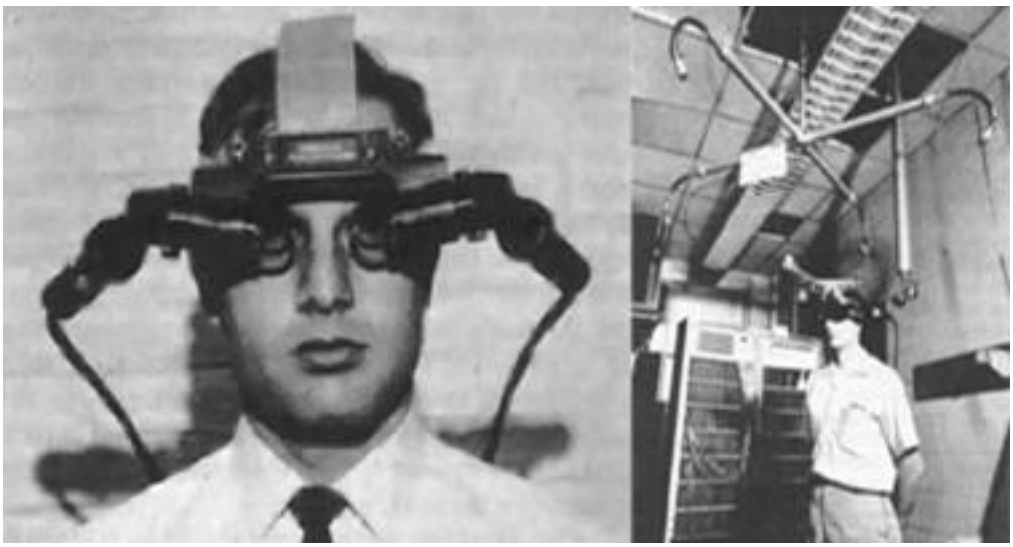


Figure 5: The Sword of Damocles (82)

by physical reality, using visual and haptic feedback (82): "*A display connected to a digital computer gives us a chance to gain familiarity with concepts not realisable in the physical world*". Sutherland went on to publish 'A Head-Mounted Three-Dimensional Display' in 1968 which described the first viable head-mounted display, this augmented reality head mounted display was able to provided stereoscopic visual images that tracked head movement. The display used miniature cathode ray tubes (similar to those used in televisions at the time) and was used to display transparent wire frame rooms and objects such a cyclohexane molecule (82). He initially used ultrasonic sensors to detect head position in the device known as the 'Shower Stall', which was followed by the 'Sword of Damocles' was a later version that used a mechanical head position sensor, named as it was suspended from the ceiling and had six degrees of freedom.

From this starting point, further development in the field was led primarily by the US Army and NASA, in their search for vehicle simulators. In the 1980s, NASA's Aerospace Human Factors Research Division developed the Virtual Interface Environment Workstation (VIEW) which was a major step forward. It was a head-mounted stereoscopic display system which allowed for input from the users hands through Datagloves, allowing for users to grasp objects

Figure 6: NASA's VIEW system (85)

(102). This work was continued by VPL, through NASA grants which lead to the creation of the first Eyephone. Despite these advances, these systems were not widely adopted. The technological limitations meant it could only generate five to six frames per second (vs 30 frames/sec generated by televisions at that time) which lead to simulator sickness. This is an important barrier to the adoption of virtual reality, and will be explored in more detail later in the thesis. In addition, the system cost $250,000, including the headsets and computers required for it to run (103). The founder of VPL research, Jaron Lanier is credited with coining the term 'virtual reality'.

Figure 7: The world's first Eyephone! (86)

The 1990s saw the release of the first mass-market standalone VR games console, the 'Virtual Boy', released in 1994 by Nintendo and used 3D stereoscopic graphics. Despite the hype around consumer virtual reality, the limited technology, lack of colour and lack of support led to commercial failure. These failures meant that the VR phenomenon was short lived, and remained limited to video games arcades rather than in the hands of consumers (104).

Figure 8: Arcade type Virtual Reality games (87)

The preceding decades of commercial failure for VR technology changed in 2010 when an 18 year-old entrepreneur, Palmer Luckey, developed the first prototype of the Oculus Rift, raising $2.4 million on Kickstarter campaign in 2012 before being sold to Facebook in 2014 for $2 billion (105).

Following this, the VR resurgence gained momentum with large technology companies such as Sony, Google, Samsung, Microsoft and Facebook developing and releasing their own systems, ranging from the cheap low-tech Google Cardboard to the HTC Vive (high-tech requiring a separate computer to run).

1.5.4 Current applications in healthcare

The fully immersive properties can elicit normal physiological responses using VR stressors (106). Currently established therapeutic targets include:

1. De-sensitisation therapy for anxiety disorders (107,108);
2. Distraction therapy to reduce analgesic requirements (109,110);
3. Patient preparation for surgery (111).
4. Assist in learning for neurologically diverse individuals (112,113)

VR has been shown to potentially improve outcomes in neurologically diverse individuals, including in conditions such as dyslexia and autism. In autism, the perceived benefits are the ability to introduce the participant to different social situations, which can reduce the stress and anxiety of real-world interactions(112). In dyslexia, the technology allows the learner's the ability to repeat and learn at their own pace, and the interactivity motivates increased learning(113).

Despite advances in medicine, the adoption within healthcare training has been limited, despite the potential perceived strengths in cueing stimuli to provide error-free learning (114). As this is an emerging field in medicine, current work in this field has been limited to proof-of-concept studies. The Royal London Hospital was the first in the world to stream a 360-degree video live. A colon cancer resection operation was filmed using 360-degree cameras and streamed to trainees equipped with VR headsets around the world (80).
Another project "We are Alfred", used stories displayed in immersive VR to enhance understanding and empathy of the aging process (115).

The first study using immersive VR for clinical decision-making was by Harrington et al. who completed a feasibility and proof-of-concept study (97). They developed an interactive training simulator using computerised VR using the Oculus platform. The study demonstrated

construct validity with regard to decision-making scores, as well as positive scores for immersion and enjoyment.

One of the disadvantages of using computerised virtual reality environments is the significant cost of production. Huang et al. showed that the development costs of a single interactive virtual patient was more than $10,000 per patient in 85% of cases and more than $50,000 for 34% (116). They went on to show that the average production time was 16.6 months, with the majority requiring more than six months to produce. Due to the inventory reporting nature of the study, they were unable to ascertain if the costs were predominantly associated with personnel or technical infrastructure.

1.5.5 360-degree immersive video

The final technology that has become available to consumers is the 360-degree camera technology. This technology offers the ability to create an immersive 360-degree recorded view, thereby– increasing the realism of the encounter. Whilst immersive VR increases alertness and immersion, 360 cameras have the potential to increase realism, as you can record real patient or actors, versus computer-generated images. This has the potential to increase transference, as the simulation is more closely related to real-life practice. In addition, the creation of cases using 360-degree cameras can drastically reduce the cost of case, which currently stands at $10,000 per patient, mainly due to the significant cost of creating computer generated patients. This technology will be discussed in more detail in Chapter 2.

1.6 Bringing it together

A fundamental principle of good pedagogy is to expose learners to examples earlier, thus allowing them to develop a database of cases from which they can build and further develop their own non-analytical processes (45). Learning from cases should be context specific and learners should be exposed to a large enough database through which they can engage in problem-solving, whilst being exposed to a wide range of presentations of the same conditions. It is also important to recognise the importance and role of sequencing.

In the real world, the ideal set of signs and symptoms may never present, whereas in a simulated world, the user can be exposed to increasingly complex presentations of the same disease, or similar presentations of different diseases. In addition, Eva has recognised the need for case progression to mimic real life application. It is therefore pedagogically optimal to expose learners to cases where they are unaware of the diagnosis at the beginning (such as those provided in a textbook whereby chapters are organised by diagnosis), as this allows development of the critical hypothesis testing phase (45).

A significant challenge to delivering simulation is to provide a realistic environment, which allows the teacher to observe the actions of the learner, in a cost-effective manner.

Immersive virtual reality designed with the principles of virtual patients using a problem-solving approach, presents the opportunity to create realistic immersive clinical scenarios for the teaching and assessment of clinical reasoning.

As with any new technology, virtual reality should not be seen as a magic bullet. Whilst simulated learning may not be as effective as learning in the real setting with real patients, due to the challenges highlighted earlier, this is no longer an option, and it is unlikely that we

will return to this state (3). The strategy should therefore be to create environments where learners can work up a series of cases that represent a specific domain.

1.7 Conclusions

This chapter began by describing the current state of medical education. It went on to describe the challenges faced, in particular, in teaching clinical reasoning and highlighted the importance of this skill in clinical practice. Having examined the history of clinical reasoning research, the current theoretical constructs and teaching methods were presented. Virtual patients can be seen as the area of most interest in furthering this skill. To this end, the latest developments in technology, namely immersive virtual reality and 360-degree cameras were presented as a technological solution that could improve upon what virtual patients are and do at present. It then went on to describe the evolution of virtual reality technologies and the relevant evidence for their application in medicine.

As reported by Turso-Finnich et al. no studies thus far have compared virtual reality head-mounted displays to another learning intervention(96). The studies to date have not been constructed in line with the existing guidelines for virtual patient creation (117). None of the published studies have approached the creation of their respective simulators in a methodological way, including the assessment of the usability of the simulator and assessing if there is an increase in presence provided by the technology, which is one of the fundamental theoretical advantages to the technology.

The strengths and weaknesses of the published studies to date have been presented in table 1. Only one of the three studies used 360 degree video, which, as presented in 1.5.5, should

increase realism, and thereby lead to more realistic behaviour and a reduction in the transfer distance when applying the acquired skills to real life.

Having established that immersive virtual reality using 360-degree videos is a potential solution to clinical reasoning training, we will move forward to Chapter 2, where the feasibility of construction is considered.

1.8 Thesis – Aims and Objectives

1.8.1 Hypothesis of thesis

We hypothesise that immersive virtual reality clinical cases using immersive head-mounted displays with 360-degree videos will allow for the teaching and assessment of clinical reasoning.

1.8.2 Aims of thesis

The main aim of the thesis is tocreate an immersive virtual reality simulator using 360-degree camera technology to teach and assess clinical reasoning skills.

1.8.3 Objectives of thesis

2. 1To demonstrate feasibility of construction of an immersive virtual reality clinical scenario using 360-degree video to teach clinical reasoning (non-experimental study)
3. 2. Construct a series of clinical scenarios to allow for further studies, (non-experimental study)Develop a scoring system to assess userperformance whilst using the simulator (non-experimental study);

3. To assess the usability of a pilot scenario across a range of potential future users – using a validated usability assessment questionnaire (Cohort study);

4. To assess if increased immersion results in increased presence (using a validated presence questionnaire)– this will assess two different delivery modes of the simulator, a computer screen versus a head-mounted display (Randomised controlled trial);

4. To examine whether the simulator differentiates across users of different clinical ability (Cohort study); and
5. To assess the learning gains in clinical reasoning following use of the simulator versus the case-based teaching method currently employed in post-graduate training (Randomised controlled trial).

Chapter 2

The development of the first immersive VR clinical scenario using 360-degree video, followed by creation of three further modules, and an objective scoring system

Introduction: In this chapter I will outline the steps taken to create the immersive 360 degree video simulator. This includes technological considerations, design principles and the principles used for the clinical case creation. I will then go on to describe the creation of a further three simulator cases as well as the development of a scoring system for their use. The first simulator was assessed for usability (see Chapter 3), the results and information gained from this study guided the creation of the further three cases and the scoring system. The three further modules and scoring system were then assessed in the experimental chapters (cohort study – see Chapter 4, and randomised control trial – see Chapter 5).

Objectives:

1. To demonstrate feasibility of construction of an immersive virtual reality clinical scenario using 360-degree video to teach clinical reasoning (non-experimental study)

2. Construct a series of clinical scenarios to allow for further studies, (non-experimental study)
3. Develop a scoring system to assess userperformance whilst using the simulator (non-experimental study);

The three phases of the clinical simulator creation have been presented in stages – technological considerations, design principles and case creation (see figure 9). Firstly, under technological considerations, various technological solutions, including camera and software platform will be presented. Following this, a summary of the design principles necessary for use in the virtual reality world are presented, these principles will be applied with the main aim of establishing familiarity with the simulator and avoiding simulator sickness. The above design principles will be applied to structure a clinical scenario. As presented in the introductory chapter, our proposed simulator would be similar to the virtual patients in the literature, therefore, the guidelines by Posel et al. were followed during this step, which included storyboarding and creation of a case on paper (117). The final step was the recording of the clinical case, followed by post-production editing, and finally the case was put together using the software platform. The layout of the clinical scenario and interactions was based on the design parameters of the first step. Whilst this final step is presented as a sequential step, it did in fact require several iterative steps and testing in order to work out the optimal use including camera angle, distance, and display parameters.

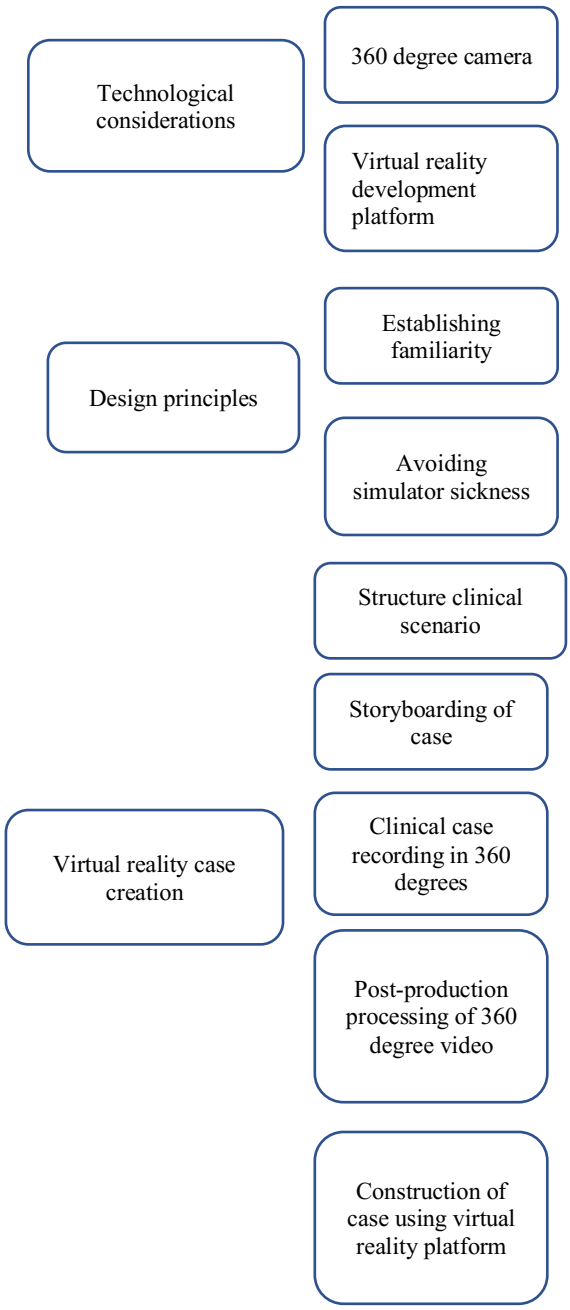


Figure 9: Consideration in the creation of a virtual reality clinical simulator

2.3.1 Technology and design considerations

2.3.1.1 Camera

Omnidirectional cameras are required to create and record 360-degree video, and can be grouped by function into dioptric, catadioptric and polydioptric (118). Conventional cameras, such as digital single lens reflex (DSLR) and phone cameras, can be described as working as a pinhole and work on the perspective projection principle (118). Omnidirectional cameras do not work using perspective projection - in order to capture a wide field of view, they are required to create very high distortion of the image being seen.

Dioptric cameras (e.g. the fisheye lens used by consumer level cameras such as the Samsung 360), use shaped lenses for refraction in order to achieve a field of view greater than 180 degrees (118). Catadioptric cameras on the other hand combine a shaped lens (dioptric) with a mirror (catoptric) (119). This gives a 360-degree FOV in the horizontal plane and 100 degrees in the vertical plane (118). Polydioptric cameras, which can produce professional quality 360-degree videos are simply ones that use several wide-angle cameras simultaneously (8-17 cameras) (119), using a rig or inbuilt into one structure, each with overlapping views. After recording, software is used to stitch together all the inputs to create a continuous and smooth view without overlap.

The disadvantage of catadioptric cameras is that they have a limited field of view in the vertical plane and are not compact (119). Polydioptric cameras, whilst offering the highest resolution are extremely expensive (starting at around £5,000). Recent advances in this technology have allowed for consumer-level 360-degree cameras which work by combining or ‘stitching’ the input from two cameras, which produces a quality image at an affordable price.

Another consideration with 360-degree cameras is monoscopic versus stereoscopic, each with its own advantages and disadvantages. Firstly, in monoscopic cameras, i.e. there is only one lens recording one point of view, therefore the image presented to each eye is identical, thereby reducing the sensation of depth (120) that can be created with more complex and expensive 3D camera. In addition, this limits the degrees of movement to three, as it allows for rotation and no translation (120). In addition, these limitations can increase the vestibule-ocular mismatch and increase motion sickness (120). Stereoscopic cameras on the other hand more closely represents the way we see the world. These cameras have two lenses on each side, as opposed to one per side for monoscopic, and are therefore able to create a greater sense of depth in videos. Despite these benefits, the drawback of stereoscopic cameras is the inability to view their output on a 2D device (such as a laptop), it can only be viewed on a dedicated virtual reality headset. In addition, the recording equipment and the post-production is more expensive.

Following the recording phase, software is used to digitally 'stitch' the separate footage into a spherical video. This can be viewed in immersive 3D using a head-mounted display, or on a normal screen in 2D, giving a panoramic view.

The requirements for the camera were:

1. Cost <£500
2. Compatibility with virtual reality production software
3. Ability to create file size <1 gigabyte per five minutes of recording (to allow for storage and post-production)
4. Ability to function and synchronise with a smart phone during recording (to allow for adjustment of camera angle, height and distance)

Based on the above requirements, the Samsung 360-degree dioptric camera was chosen. It is worth noting that due to the price constraints, this camera has a resolution of less than 4K, and is therefore likely to show less high-quality resolution than the higher-end cameras such as the GoPro Odyssey, which cost around $15,000 for the system. The lower resolution quality may affect overall usability, but this will be assessed in detail in the following chapter.

2.3.1.2 Software platform

Due to the novelty of this technology and the specific use case we are defining i.e. to create a clinical decision-making simulator, there is no software that currently exists to create the complexity of interaction required. Instead, several companies were contacted and discussions had in relation to the way in which to retro-fit the software for our needs. Current commercially available technology is designed for simple interactions, however, our aim was to recreate a clinical scenario, requiring up to fifty interactions per screen. After testing multiple platforms, a combination of Wonda VR and Viar360 were used for case creation. WondaVR is a desktop-based 360-degree editing programme whilst Viar360 is predominantly web-based.

2.3.2 VR design principles

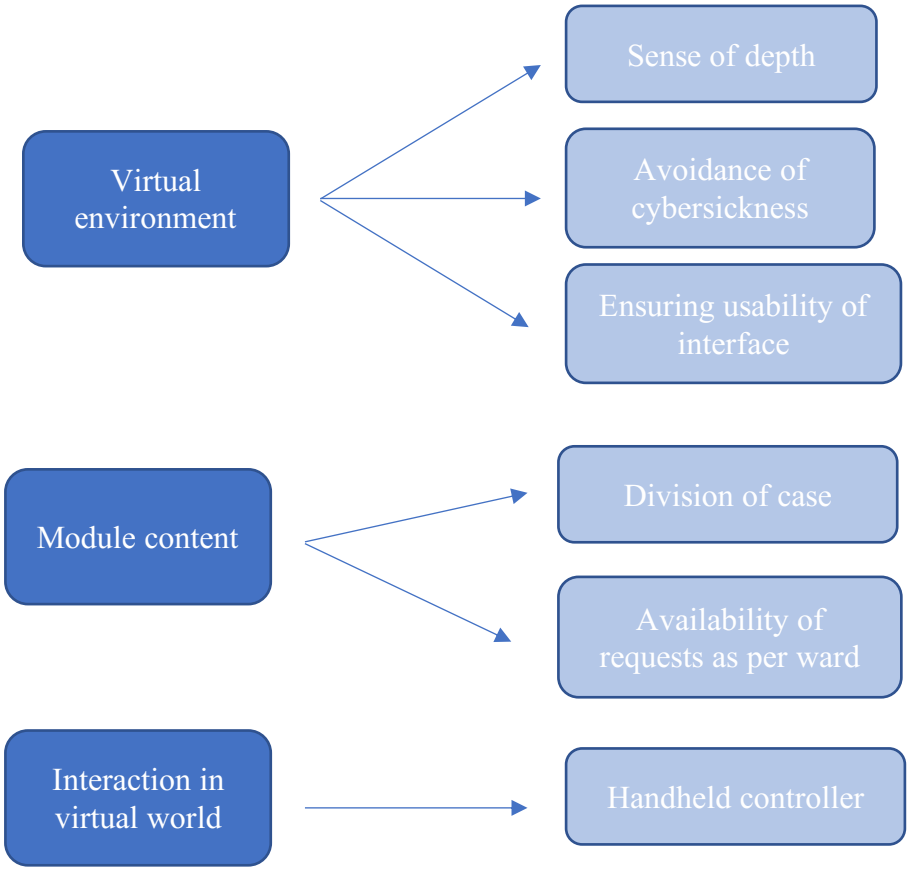


Figure 10: Key elements for the clinical decision-making simulator

Due the novelty of the technology, as well as context-specific challenges, designing a user experience in virtual reality is much more difficult than 2D, for reasons I will present shortly. I used the guidelines created by the Google Cardboard team in 2016 to aid the creation of the virtual clinical scenario (121). The two main themes of the guidelines are:

I. Avoiding simulator sickness; and

II. Establishing familiarity.

Some basic terms to be established include 'roll', 'pitch' and 'yaw'. 'Roll' is a movement whereby the user's head pivots from side to side in the coronal plane. 'Pitch' is a movement whereby the user's head moves up and down in the sagittal plane. 'Yaw' is a movement whereby the user's head rotates in the axial plane.

2.3.2.1 Avoiding simulator sickness

Using immersive virtual reality can lead to mismatches between physical and visual cues, resulting in motion or simulator sickness. This is a known and unfortunate complication of this technology which has greatly reduced due to improvements in the technology, which has improved latency rates for the screens. There are still some users who suffer from simulator sickness and therefore strategies have been developed to reduce these (121,122):

a. *Head tracking* – this allows objects in the virtual world to maintain their position, regardless of how you move your head. This is essential as it reduces the mismatch between the visual system and the vestibular system, as even slight pauses in head tracking can cause users to feel sick.
b. *Avoiding slight freezes in head tracking* – it is important to pay close attention to these when testing the experience. If this occurs during transitions, a solution is to fade the screen to black to maintain tracking.
c. *Rendering 2D splash screens in a 3D space* – when rendering 2D splash screens, it is important to ensure that head tracking is maintained. Three degrees of freedom (rotation, pitch and yaw) are preferable to one (rotation) in reducing nausea.
d. *User controls movement* – users should not be 'passengers' with respect to movement, i.e. they should maintain control of their movement inside the app. This allows them to anticipate movements thereby reducing nausea. For instance, a simple video recording showing a point-of-view (POV) shot is more likely to cause nausea than one where the user controls all the movements.

e. *Use a constant velocity* – an acceleration or deceleration within the experience will not be felt by the user (this dissonance can lead to nausea). It is important therefore, that if there is movement, it is at a constant velocity.

f. *Grounding with fixed objects* – if the user is in a fixed virtual environment, they can be seated in a chair to reconcile the fact that they are not moving in real life. If the user is near a large moving object, they may feel they are moving, and it is therefore useful in such cases to have more fixed reference points in the user's environment.

2.3.2.2 Establishing familiarity

Most people have not experienced immersive virtual reality before. As the VR canvas is so large, a particular challenge is guiding the user's focus and attention. The relevant points to be considered in the construction of the simulator are presented below(121,122):

a. *Set-up* - getting situated with the virtual reality set-up can take time. In order to avoid the user feeling rushed, ensure that the experience only starts once the user is ready to proceed, through use of a button, rather than through a timer.

b. *Head-set adaptation* - the app should be tailored to suit the physical characteristics of the headset being used.

c. *Controls* – when the experience launches, the user interface should be presented to the user's current field of view. If this is not actively updated, the user will feel confused and have to look around for the controls.

d. *Fuse buttons* – if the headset does not have an option to allow for clicking on targets, a fuse button may be used that requires prolonged gaze for action. This can be slow and frustrating and should be avoided if possible. These buttons should have visual countdowns to alert the user that an action is ongoing. Fuse buttons should work best

if they are large and far apart from one another, as this will avoid clicking on the wrong button, and inadvertently moving off the active button.

e. *Audio* – due to the visual overload in the virtual environment, text instructions do not perform as well as audio. When used, instructions should be brief.

f. *Display reticle* – users can accurately target small objects by moving their gaze. It is easier to do if there is a reticle displayed, though this can reduce the sense of immersion and can create clutter. To reduce these effects, the reticle can be displayed only when the user approaches an active button or to ensure that the buttons hover when selected, to signal that they are active.

g. *Comfort zone* – head movement should be restricted to the limitations of the body. Whilst this is a rather obvious point, it translates into a user interface canvass of 94 degrees in the horizontal plane and 32 degrees in the vertical plane, for a person sitting in a fixed chair.

It is important to note that the guidelines and principles used were borrowed from the relatively new field of virtual reality design. These virtual reality principles were developed and designed for simpler use cases than the ones proposed by this thesis.

For example, many layout suggestions are simply to allow users to navigate through a library of movies, or pictures, rather than to interpret complex data with multiple options. The number of options and decision-making parameters is also far greater than what would be expected for the technology in its current use. The potential result is that the layout may be deemed unsatisfactory both in terms of its presentation of a clinical case, and in terms of the number of options available to the candidate. The usability of the simulator produced will be

formally assessed in subsequent studies using validated tools such as the System Usability Scale and Face Validity Scale.

2.3.3 Structure of a case

The move to immersive virtual reality can be seen as a shift along the continuum from the virtual patient to the real patient. It seems most appropriate to therefore base the structure of an immersive virtual patient on both.

The basic structure of any real-life clinical encounter begins with taking a history to establish differentials, followed by an examination, informed by the history, followed by simple/bedside investigations, depending on the results of these, further investigations and actions can be taken.

With the creation of virtual patient cases, Posel et al. described guidelines for their construction, some of which apply in this situation (117).

Relevant guidelines include:

1. *Determine case content and choose a design model*

In this case, whilst it is a new technology, it can be described as a branching case, i.e. one that allows for the highest level of interactivity for the user, whereby the order of interaction is not pre-determined and the consequences of which are different based on decisions made.

2. *Organise and storyboard the case before starting*

Storyboarding of the case was carried out on Microsoft Powerpoint and divided up as presented above. All data sets were completed and agreed among the authors prior to recording.

3. *Intuitive and logical site navigation*

   The rules of VR design and user interface design were applied to create an intuitive immersive clinical environment. The acceptability of this will be assessed in Chapter 3 using the System Usability Scale.

4. *Tackle interactivity*

   The entire premise of immersive virtual reality is to increase immersion and interactivity by increasing the presence for the user within the virtual world. This is further enhanced by the use of a handheld controller versus a head mounted control, which inherently reduces the complexity of interactions that can occur.

5. *Post-case evaluation*

   This will be carried out as part of the usability test in Chapter 3.

6. *Use of expert traces/script concordance*

   Expert traces and script concordance are methods both for showing the clinical reasoning strategies of experts as well as for assessing the clinical reasoning of users. This idea is further explored in Chapter 6.

7. *Choose the right authoring application*

   Unlike the virtual patient authoring tools, which are ubiquitous, there are only a handful of immersive VR authoring platforms, as this is an emerging technology. The chosen platform is described at the start of this chapter.

2.3.3.1 Case selection

Post-operative pyrexia was chosen as the initial case as this is a clinical presentation that is commonly seen on a surgical ward and for which early diagnosis and appropriate management is essential.

2.3.4 Case creation

2.3.4.1 Storyboarding the case

A case was created in line with the guidelines described above, in collaboration with two nationally recognised surgical trainers. The structure of the case is outlined below:

1. Handover from multi-disciplinary team member
2. Patient records
3. Patient history
4. Examination
5. Simple investigations
6. Further investigations
7. Intervention

2.3.4.2 Filming

Filming for the case was done over two days using the Samsung 360-degree camera. All the responses were scripted in the storyboarding phase. A second-year medical student played the part of the patient, with a third-year medical student playing the role of the nurse.

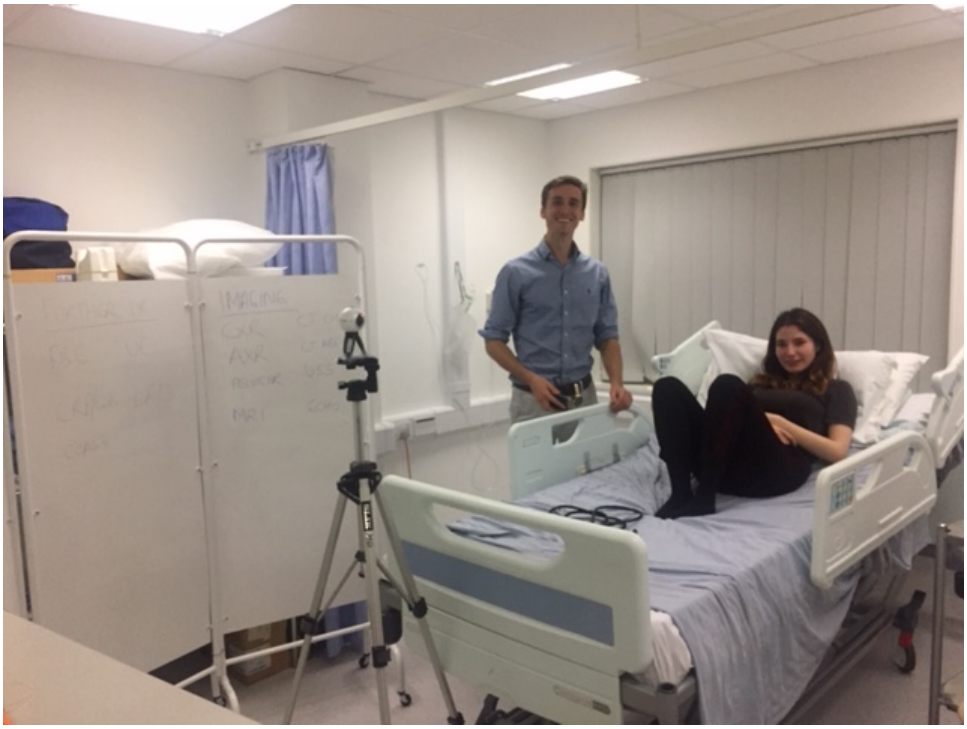


Figure 11: Showing the set up for recording, with the 360 degree camera placed at the foot of the patient bed

2.3.4.3 Post-production

Following filming, all the videos were imported into the software platform for post-production editing. This phase involved linking the relevant videos to one another, in addition to inserting corresponding relevant images (e.g. observation charts), to the relevant buttons. This process took around six months. the menu options were created on the software platforms and structured as described in the structure of a case (2.3.4.1). Please see appendix 1 for the raw tables used to display the patient results. Figures 15-19 show the interactive screens created for navigation through the case.

2.3.4.3.1 Basic user testing

Following post-production, the module was published to the application as a beta version.

User testing was carried out to ensure identify and resolve any bugs in the simulator.

Figure 12: Showing the Samsung 360 camera placed on a tripod to allow an unobstructed view whilst filming

Figure 13: Showing the Pixel 2 phone, Google Daydream headset with accompanying controller

Figure 14: Pictures showing the testing for bugs using the headset with controller

2.4.1 Pilot case summary

Asked to see Ms. Smith as she is pyrexial, three days post-op for an elective ileo-caecal resection for failure of medical treatment. She has an acute worsening of abdominal pain, associated with vomiting and pyrexia.

The user is expected to go through the admission clerking, operation note (which suggests a difficult operation due to failure of the staple gun which necessitated hand sown suturing of the anastomoses), and all available blood and further investigations to reach a decision. In this case, this young lady has had an anastomotic leak which would necessitate i.v. antibiotics and preparation for theatre.

Presented below are a series of screenshots from the clinical scenario.

Figure 15: Screenshot of the clinical reasoning simulator

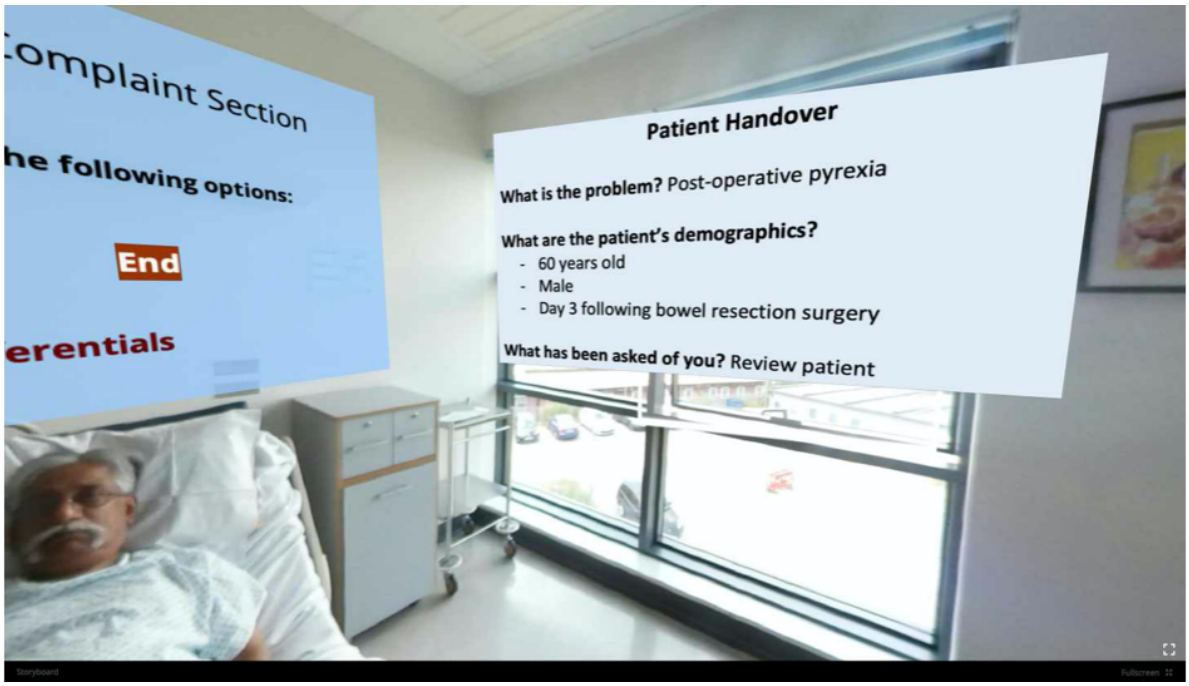


Figure 16: Screenshot of the clinical reasoning simulator showing the patient handover

Figure 17: Screenshot of the clinical reasoning simulator showing the differential diagnosis list

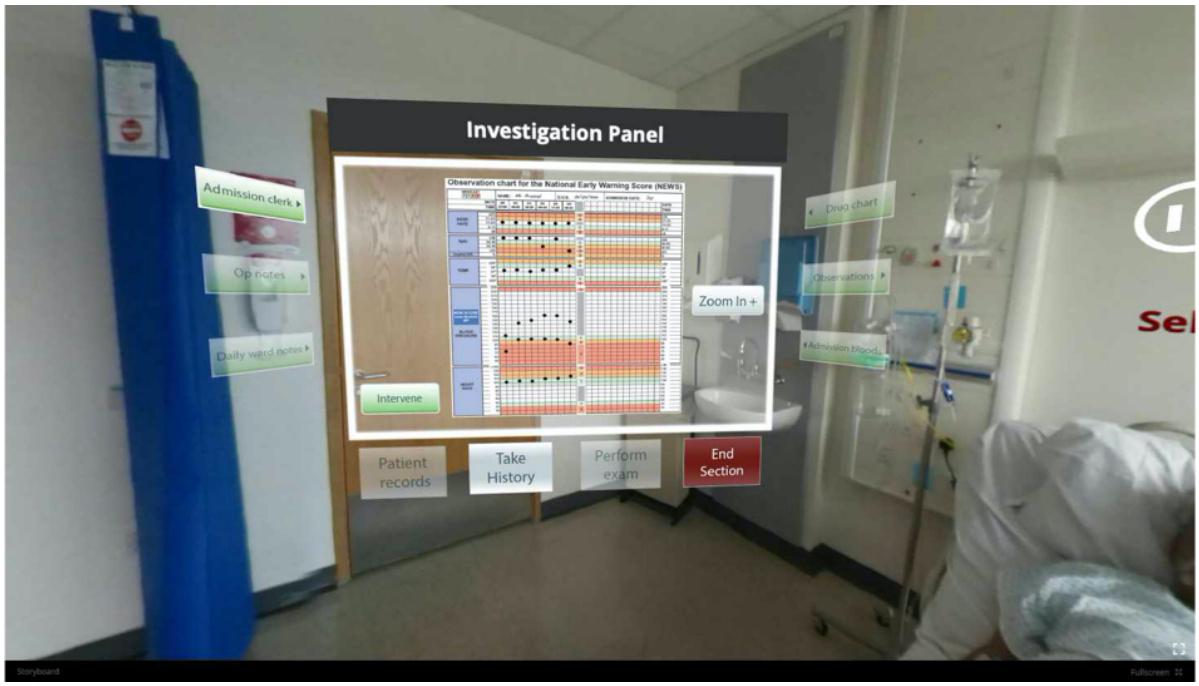


Figure 18: Screenshot of the clinical reasoning simulator showing the investigation panel

Figure 19: Screenshot of the clinical reasoning simulator showing the investigation panel

This study was successful in achieving its end goal of a working clinical scenario which could later be tested for important characteristics such as usability, cybersickness and presence, which will all be essential in validating its further use.

As this field is relatively new, and the use case is the most complex of the those being carried out, there is a limitation on the level of interactivity that can be provided by this system. The aims of the initial design brief were achieved, however, some more aspirational aspects of simulation were limited by the software available.

The ability to have an immediate change in patient parameters based on treatments given. The current ability of the system will allow for changes to the patient to occur after each stage is completed. In addition, the patient can undergo a series of changes or deteriorate within the same stage. It is not possible for the physiology of the patient to change in response to each individual management option administered, e.g. if a candidate chooses to give oxygen,

it will not suddenly appear on the patient, it will however appear on the patient as the candidate proceeds to the next stage. This will likely have an effect on the perceived 'realism' of the clinical encounter and will therefore be picked up on the face validity questionnaire.

In chapter 3, the usability of the simulator and the effectiveness of the immersive delivery method on creating presence will be presented. The results and learning from that study were used to create three further clinical cases and a scoring system for use in further studies, which is presented below. 2.7 The creation of the three further cases

Three cases were created using the same methodology as the first simulator case. This process included story-boarding, recording of the cases, post-production, construction on the software platform and de-bugging. The topic for the cases chosen was post-operative sepsis, as is a leading cause of multiple organ dysfunction and in-hospital mortality (123). It is something that has shown variability in management, and is a fairly complex process, based on the knowledge of differentials, a systematic approach, appropriate interpretation of the history and notes, appropriate investigations and early and correct management. These characteristics make it ideal for the teaching and assessing clinical reasoning. In addition, it has been demonstrated that the ability rescue a patient from a complication is both to do with timely pickup and appropriate management (124).

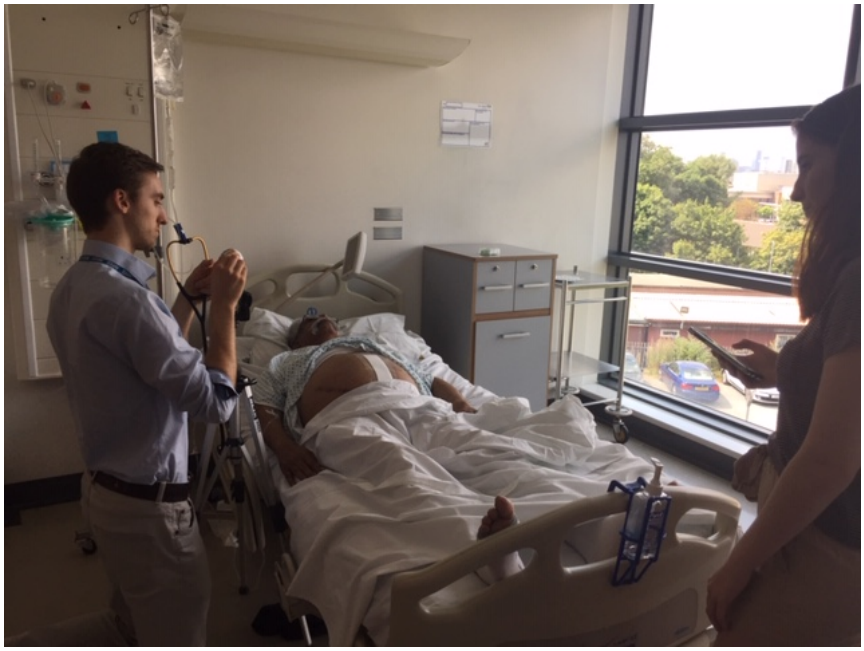


Figure 20: The recording of case 1

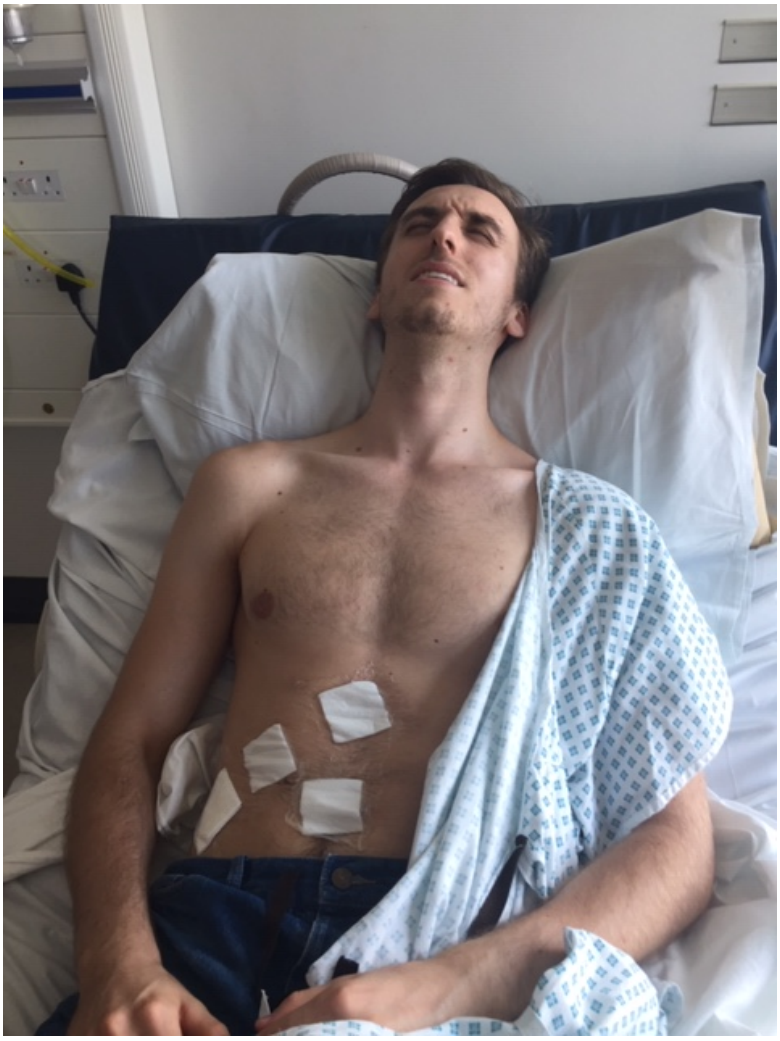


Figure 21: The patient actor in case 2

Figure 22: The patient actor and camera set-up in case 3

2.7.2 Immersive case scenario 1

Description of Mr. Prasad (full details in Appendix)

Mr. Prasad is a 65 year old gentleman who is three days post emergency right hemic-colectomy. The candidate has been called to see the patient as he is pyrexial. He was admitted following 48 hours of abdominal pain and distension, on a background of anorexia, constipation and lethargy for the preceding weeks and months. He has multiple co-morbidities including unstable angina, for which he is due to undergo an elective coronary angiogram (this make a differential of acute MI a possibility for a fever). Following admission, he underwent an emergency open right hemi-colectomy, which was uneventful. In the immediate and early post-operative period he had a high NG output, which was pulled out by the patient on day 2. In addition, he was started on a PCA due to difficulties with analgesia. The information provided suggests the patient has ileus, poor pain control which is causing a distended abdomen post-op with a splinted diaphragm. Blood results reflect an infection,

which the examination shows is likely pneumonia, confirmed with a chest radiograph. The most appropriate steps in management include oxygen, antibiotics, re-insertion of the NG tube, fluids and more close monitoring.

2.7.3Immersive case scenario 2

(full details in Appendix)

Ms. Wilkinson is a 25F admitted 2/7 ago with lower abdominal pain. The candidate has been called to see the patient as she has ongoing pyrexia and worsening abdominal pain. On admission, she was pyrexial, had nausea associated with one episode of vomiting, and had a 24 hour history of diarrhoea. She reports a previous episode of abdominal pain at the time of her menses, she is sexually active and nulliparous with no other gynaecological history of note. No other co-morbidities. Whilst in hospital she has spiked temperature of >38.5C, had a few episodes of vomiting and she continues to experience loose stools. She has been reviewed by the gynaecological team who have cleared her. Pelvic USS was negative for gynae path however the appendix was not visualised. The information provided suggests an inflammatory/infective process, with the abdomen as the source for either. Examination and further investigations would reveal an inflamed and perforated appendix that requires antibiotics and urgent operative intervention.

2.7.4 - Immersive case scenario 3

(full details in Appendix)

Mr. Barrett is a 24 year old medical student 2/7 post elective laparoscopic cholecystectomy for USS confirmed biliary calculi. The candidate is asked to see the patient as he is pyrexial and complaining of abdominal and shoulder tip pain. The operation was reported as

uneventful and the patient was kept in overnight as he was nauseous. His post-operative bloods are in keeping with his surgery, no unexpected findings on examination. Whilst the history, examination and investigations suggest a post-operative SIRS/basal atelectasis.

2.5.1Development of an objective performance scoring system

A key stage in the development of any simulator is the development of an objective scoring system. The immersive virtual reality simulator developed can be seen, in terms of the literature, between existing virtual patient systems and high-fidelity simulation. Therefore, the scoring system is based upon the existing literature in the field of virtual patients. Assessment of virtual patients have been defined by Round et al. as 'computer based simulations of patient management designed to predict performance in clinical setting' (125). The advantage of virtual patients for assessment include standardisation/reliability, as all candidate are assessing the same patient, high face and predictive validity as it combines different media in a way that a paper-based examination cannot (125). These advantages have led to their widespread adoption such as in the US medical licensing examination, where it has been shown to assess a different proficiency to that assessed by MCQ's (126).

Despite its widespread use, in two national licensing examinations, the guidelines for the development of objective scoring for virtual patients is not clearly defined. A recent paper by Fors et al. compared eight different scoring rubrics in a second-year medical school cohort and described that the appropriate rubrics to be used should be based on the objective of the assessment (127). The immersive simulated cases are designed for final year and post-graduate students, and therefore require a more complex blend of rubrics which I will set out below.

2.5.2Principles of scoring

2.5.2.1 Division of inquiries

Any interaction that is made with the simulator is called an inquiry, i.e., if the user clicks on the button to perform a respiratory examination, it would be registered as an inquiry. Each inquiry must then be assigned a value as to its appropriateness. Most simply, this can be done by assigning each inquiry as appropriate or inappropriate. As suggested by Fors and Gunning, a differential weighting can be applied to each inquiry (127). We assigned three levels of appropriateness to each item:

Correct – if the item is deemed to be appropriate in managing the case

Irrelevant – if the item is deemed to be unnecessary in managing the case, though not harmful to the patient e.g., ordering a urine dip for a patient with no specific indication

Incorrect – if the item is deemed to be unnecessary and potentially harmful to the patient e.g., ordering a CT abdo/pelvis when not indicated, due to the high levels of radiation; or incorrect – not keeping the patient nil by mouth when suspecting

Appropriate division of all responses should be into correct, irrelevant or incorrect

2.5.2.2Controlling for shotgun approach

A key part of a scoring system will be controlling of a shotgun approach. This is defined as when a novice, unsure of what is correct, simply chooses all the options available in an arbitrary manner. In doing so, they demonstrate high sensitivity for the correct options, but low specificity for the incorrect options. In order to control for this, the number of correct

options should be weighted against the total number of options selected, thereby penalising a user if they randomly select options.

In keeping with the literature, the final score will therefore be calculated by the number of correct enquiries minus the incorrect enquiries, divided by the total number. This therefore scores appropriately correct answers, penalises incorrect answers and controls for the 'shotgun' approach, which would be expected by the less experienced learners:

Total score for clinical scenario = (Number of correct enquiries – incorrect enquiries) / Total number of enquiries

2.6.1 Domains of scoring

The simulator is designed to assess clinical reasoning, therefore the assessment will be broken down into:

- History
- Differential generation
- Examination
- Simple investigation
- Further investigation
- Simple management
- Further management

Combing the above, each domain will have a number of inquiries, which will each be assigned a designation of correct, irrelevant, or incorrect. Below is an example of the designation of responses for the history domain

Table 1: Designation of responses

Interaction	Designation
Ask about wound history	Correct
Ask about travel history	Irrelevant
Request bone scan	Incorrect

2.7.1 Determination of scoring system for each case

Following creation of the cases, a group of surgeons consisting of two national surgical trainers and two post-graduate MRCS surgeons discussed each inquiry within a case, and assigned it as correct, irrelevant or incorrect. A consensus was reached on all assigned actions with no disagreement between the surgeons. This was documented and formed the scoring system against which future users would be marked.

The process of creating three cases and developing scoring for each one took 12 months to complete.

2.7.5 - Scoring system

A marking scheme was created for each inquiry for all three cases. A sample of the scoring system for scenario 1 is shown below (full scoring system in Appendix).

Table 2: Sample scoring system for Scenario 1

Section	Inquiry	Agreed response
Patient records	Admission clerking	Correct
	Daily ward notes	Correct
	Operation note	Correct
	Admission bloods	Correct
	Drug chart	Correct
	Observations	Correct

The scoring system was devised by gaining consensus amongst a group of surgeons, as has been validated in the past. The aim of the system is to be in used in conjunction with the simulator, to produce a score which defines performance on the simulator for each respective case. A limitation to note is that the experts providing the agreed response for each enquiry were using the simulator on a laptop screen, it is possible that their behaviour, and therefore the agreed response may have been different in the immersive environment.

The final scoring rubric ((Total score for clinical scenario = (Number of correct enquiries – incorrect enquiries) / Total number of enquiries)) that was chosen is born out of the literature,

as it allows for the calculation of correct answers whilst also penalising incorrect ones. In addition, by dividing by the total number of enquiries, it takes into account the Shotgun approach that may be used by juniors. As has been noted in the literature, there are advantages and disadvantages to every scoring rubric. The one devised for this thesis is believed to be the best for the target audience of senior medical student and junior doctor, but may be less effective when applied to expert clinicians.

2.8 Conclusion

This chapter presented the creation of the first clinical case, followed by three further cases followed by a scoring system.. The next chapter will

Chapter 3 – Usability testing of the simulator

3.1 Introduction

Having demonstrated the feasibility of creating a clinical scenario using immersive 360-degree video, the primary aim of this study was to assess the usability of the simulator by the intended future users.The usability of a tool, or, its fitness for purpose, is an important factor when creating a new tool. Whilst usability does not exist in an absolute sense, and is context specific, there is agreement in the literature that tools do exist to assess this hard to measure quality (128). As such, a validated System Usability Scale was used, which is presented in the methods section.

One of the potential disadvantages of immersive virtual reality is the potential to cause motion sickness due to the vestibular-visual mismatch, which is termed 'cybersickness'. The symptoms mirror those of classic motion sickness, though the user is stationary (129). A long list of symptoms have been described including: eye strain and headaches on the milder end and ataxia, nausea and vomiting on the more severe end (129). Whilst similar in presentation, the mechanisms underlying motion sickness and cybersickness are different. Vestibular stimulation alone can cause motion sickness, whilst cybersickness occurs only with visual stimulation (129). Currently known strategies for reducing this were employed in the design phase, however, it is a known negative effect in virtual environments. It was hypothesised that the treatment group would have slightly higher scores than the control group, though it was felt that the strategies employed to reduce the effects should result in a small difference.

As discussed in the first chapter, when assessing any simulator, one of the basic validation tests is face validity, i.e. the degree of semblance between the simulator and that which it is simulating (36) (130). As the simulated case was created using 360-degree video, the face validity is expected to be high for both groups. The immersive nature of the simulator would likely result in a higher score for the treatment group.

As discussed in the introductory chapter, one of the main advantages of an immersive virtual environment is the creation of presence. Whilst it may seem obvious that wearing a head-mounted display will result in increased presence versus looking at a desktop computer screen, this is not necessarily the case, as the subjective experience of presence is context-dependent. The effect on presence is also the main reason for using immersive virtual reality, versus designing the same simulator in 2D.

The effect of immersive technologies in the literature has been mixed, as it has been shown that the newer technologies may instead distract users, increase sickness and lead to reduced learning outcomes and enjoyment (85). Due to the limitations of the technology at this stage, it may provide an experience that turns people away from VR as a medium (85). Presence will be measured using a validated questionnaire which is presented in the 'Methods' section. It is hypothesised that as the head-mounted display allows the user to be completely immersed, as well as to move with 3 degrees of freedom, that the treatment group would have higher scores for presence.

Hypothesis: **Usability – participants in the control group should have higher usability scores than the treatment group**

It was hypothesised that due to the increased complexity of using a new technology, interacting with menus in a novel way, i.e. through a clicker whilst wearing a head-mounted display, as well as presenting data and options in a 360-degree environment, that the treatment group would have lower usability scores than the control group.

Aim – the primary aim of this study was to assess the usability of the simulator across two different delivery modes, the immersive virtual reality (treatment) and computer screen (control) delivery modes.

Secondary aims – to assess different characteristics of the simulator (Cybersickness, face validity, presence)

Specific objectives:

1. To compare the usability of the simulator across two delivery methods via a validated questionnaire.
2. **To compare the level of cybersickness in participants across both delivery modes**
3. **To compare if the participants think the simulator represents real life (Face Validity) across both delivery modes.**
4. **To compare the reported level of presence by participants across both delivery modes**

3.4 Methods

3.4.1 Study design

This was a prospective randomised controlled trial comparing a clinical simulation scenario on a 2D screen (control group) versus immersive virtual reality (treatment group).

3.4.1.1 Setting

This study was undertaken as part of the simulation teaching programme at a National Surgical Conference over two days.

3.4.1.2 Randomisation

Following consent, all participants were randomised using a random number generator, into the control and treatment group. If a group was filled, all subsequent participants were automatically placed into the other group until the total numbers were reached.

3.4.1.3 Subjects

A range was chosen to ensure that this simulator was suitable for both undergraduate and postgraduate use. The inclusion criteria was any level of medical student and any level of trainee. Exclusion criteria included consultants and anyone with a previous histoy of

cybersickness with use of virtual reality simulation. Discussion amongst the research team and comparison with similar simulator studies showed a total of 15 participants in each arm would be sufficient.

In total, 16 medical students, 13 foundation/senior house officers and 10 specialist registrars took part in this study, and these participants were divided into two groups with 19 in the control group and 20 in the control group.

3.4.2 Measurement instruments

3.4.2.1 System Usability Scale

The System Usability Scale (SUS) is a ten-item Likert-based scale that has been shown to be robust and reliable in assessing subjective experiences of usability (128). To calculate the SUS score, each item has a score of 0 to 4. For the odd-numbered questions (i.e. 1, 3, 5, 7, 9) the score contribution is the mark given minus 1. For the even-numbered questions the score contribution is 5 minus the mark given. This results in a potential total of 40, which is multiplied by 2.5 to form a score out of 100 (76).

3.4.2.2 Cybersickness questionnaire

A validated questionnaire that assesses if and to what degree the participant suffered from a range of 16 symptoms. The validated questionnaire clusters the three distinct symptoms groups which reflects the effect of the simulator on each specific target system: oculomotor (eyestrain, difficulty focussing, blurred vision, headache), disorientation (dizziness, vertigo), nausea (nausea, stomach awareness, increased salivation, burping) (131).

3.4.2.3 Face validity

A Likert-style questionnaire was developed to assess the face validity of five main components of the simulator, namely history taking, examination, investigation, management and the overall representation of the ward (see Appendix 6).

3.4.2.4 Presence questionnaire

Presence is a powerful concept, discussed at length in the literature (89), and it has been accepted that presence questionnaires have their limitations (132). As there are currently no established alternatives to these evaluation scales for the assessment of presence, such questionnaires continue to be employed. The present study uses a set of questions derived from the SLATER-USOH-STEED questionnaire (133,134) presence questionnaires are not always able to detect a difference among different modalities of virtual interaction. In this study, we used the presence questionnaire to explore if experiencing the same training system (i.e. environment) in a 3D or a 2D modality resulted in a different level of presence experienced by participants. The full Presence questionnaire can be found in the appendix.

3.4.3 Equipment

Control group – this was administered using two Apple Macbook laptops, all participants used headphones.

Treatment group – this was administered using a Google Daydream 2 head-mounted display and clicker, with a Pixel 2 phone. All participants used headphones.

3.4.4 Study process

Following randomisation, participants were oriented to the appropriate simulator by the lead researcher. This began with a demonstration of the case on the laptop, showing all the menu options to ensure candidates felt comfortable with the layout.

Following this, the treatment group were oriented to the use of the headset and completed an introductory module to ensure they were familiar with the simulator. The control group completed the same introductory module on the laptop. The System Usability scale was completed by both groups after completion of the introductory module. Participants then started the actual clinical scenario and were allowed as much time as needed for this. Following completion of the module, candidates completed the remaining questionnaires. In addition, informal feedback was received from the trainees on their thoughts of the learning experience.

3.4.5 Statistical analysis

All data was transcribed from the written survey onto Excel and subsequently imported into Stata v 13.0. The unpaired T test was used to determine significance in means between the control and treatment groups.

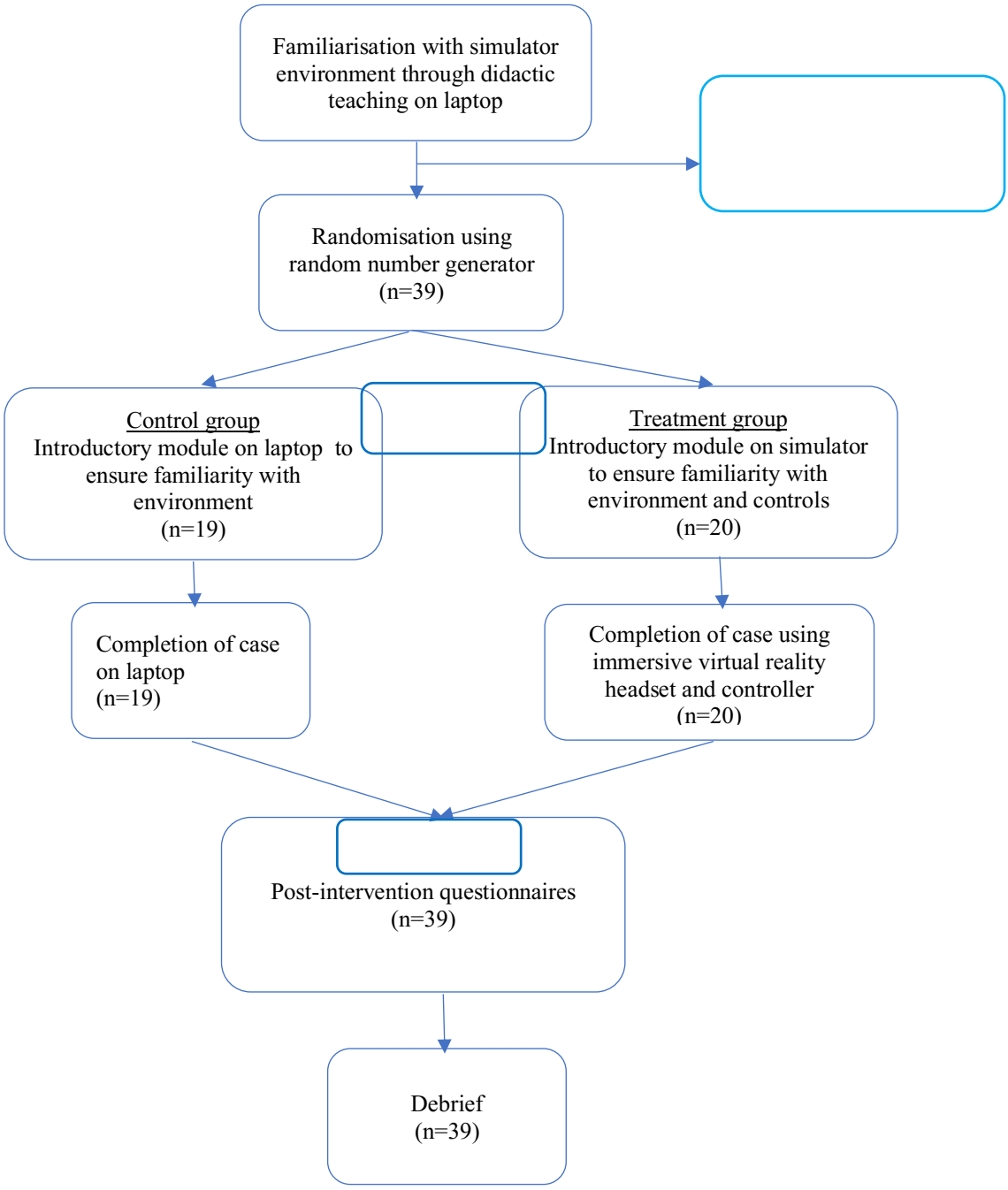


Figure 23: Study process

3.5 Results

39 participant took part in the study, the breakdown of the control and treatment group are shown in Table 3.1. In total there were 19 participants in the control group and 20 in the treatment group. All participants completed the study process outline in figure 20.

Table 3: Distribution of participants by training level

	Control group	Treatment group	Total	Male:Female	Average age (range)
Medical student	8	8	16	8:8	22.5 (19-25)
Senior house officer	6	7	13	6:7	26.1 (23-32)
Specialty registrar	5	5	10	4:6	33.7 (29-50)
Total	19	20	39		

3.5.1 Primary objective – Usability

Table 4: Usability scores

Usability	Control	Treatment	P value
Combined	**80.26**	**80**	**0.472**

The results show no difference in the usability between the immersive virtual reality simulation system and the 2D laptop simulation. Sub-group analysis showed no statistically difference in presence between males and females in the control and treatment groups ((Control group 80 vs 80.6 (p=0.45), treatment group 82.25 vs 77.5 (p=0.22)). Sub-group analysis also showed no difference in scores between people who had previously used VR and those who had not, though the usefulness of this is limited as the number who had previously used VR was 2 for control group and 1 for the treatment group.

3.5.2 Secondary objective – Cybersickness

Table 5: Cybersickness scores

Cybersickness	Control	Treatment	P value
Combined	**0.263**	**0.526**	**0.706**

The results show that the treatment group had a higher combined cybersickness score, though this was not statistically significant.

3.5.3 Secondary objective – Face validity

Table 6: Face validity scores

Face validity	Control	Treatment	P value
Combined	**26.368**	**26.263**	**0.5236**

The results show that there was no difference in the face validity between the immersive and 2D systems.

3.5.4 Secondary objective – Presence

Summarised results for Presence Questionnaire

Table 7: Presence scores

Presence	Control	Treatment	P value
Combined	**0.674**	**0.809**	**0.0102**

The results show a statistically significant difference in the presence scores between the treatment and control group. This suggests that the simulator increases the feeling of immersion for the user. Sub-group analysis showed no statistically significant difference in usability between males and females in the control and treatment groups ((Control group 0.69 vs 0.67 (p=0.42), Treatment group 0.80 vs 0.81 (p=0.43)).

Figure 24: Showing two participants using the simulator through the immersive headset and handheld controller

Figure 25: Showing two participants using the simulator on laptops

3.6 Discussion

This study set out to test four main hypotheses. Namely if the simulator was usable, if it cause cybersickness, if it had face validity, and finally if there was a difference in presence between the two modes of delivery (i.e. immersive with headset versus non-immersive with a laptop).

1. To compare the usability of the simulator across two delivery methods via a validated questionnaire.

–The results demonstrate that there is no significant difference in the usability scores between the two groups. This was an unexpected finding because the immersive VR tool is a much more complex environment to navigate. This suggests that the basic VR design principles have been appropriately applied in this context. The most important aspect of this

finding is that the 2D is a suitable alternative, in terms of usability, to the 3D. Though the platform has been designed for use in an immersive 360-degree environment, the simulator can be used equally well in 2D. This is particularly relevant to its use if the required head-mounted display is not available, or if the users experience significant cybersickness during its use.

Another interesting aspect which requires further study would be the relation of usability to age. The target audience for this technology is final year medical students and the first few years of post-graduate training, who are aux fait with most technologies. This generation may perhaps have a higher capacity or plasticity for learning new technology than the previous generation. An interesting study could address the SUS with the Consultant body. Whilst an interesting question, this is outside the scope of this thesis and will not be investigated further.

2. – **To compare the level of cybersickness in participants across both delivery modes**

Participants in the treatment group had higher cybersickness scores, though this was not statistically significant. As discussed earlier, one of the major drawbacks of any immersive system is that it can lead to motion sickness.

An finding important of note was that all participants were able to complete the simulation, suggesting that the degree of sickness was not severe. This is well below the reported rate in the literature of 22-80% (135). A significant determinant of cybersickness is technology-related factors, as presented earlier (136).

In addition, there were no adverse events, such as simulator induced vomiting. Whilst the difference was not statistically significant, it is possible that the sample size was too small in this case to detect a difference. This poses a question for further evaluation in the future. As described above, the usability score was the same for both systems, suggesting that it could be used in 2D for those who were intolerant of the head-mounted display.

3. **To compare if the participants think the simulator represents real life (Face Validity) across both delivery modes.**

The results show that the simulator has high face validity in both immersive and non-immersive methods. As the underlying platform was the same for both groups, i.e. menus and options, this finding is consistent with expectations. It also further validates the point that the system that has been designed can be used in both 2D and immersive VR. In a similar way to the findings for hypothesis 4, there is a possibility that this was a type II error as the participants may have relativised their scores for the system being used. Many participants commented on how impressive the system was, therefore, there is a possibility that directly comparing 2D versus 3D in separate groups, rather than asking the same group of people to compare both technologies, led to a lack of difference being noticed. This area warrants further exploration but would be out of the scope of this thesis.

4. **To compare the reported level of presence by participants across both delivery modes**

As hypothesised, participants in the treatment group had higher scores for presence than the control group. Whilst it may seem obvious that wearing a head-mounted display which responds to movement, leads to greater immersion, this does not mean that it creates a

higher level of presence (as discussed in Chapter 1). This is due to the task dependent nature of presence.

Importantly, the evidence from the literature regarding presence measurement instruments, suggests that these instruments can fail to report a difference between technologies used. This can be explained as a type II error, and has been described as occurring due to the fact that participants relativise the environment that they are in. In this study, participants would use the rating scale on a relative basis to the method of delivery they were randomised to. I.e., you are scoring the simulator used on a laptop against itself, rather than against the simulator in immersive mode. It has therefore been reported than due to the susceptibility of presence measurement instruments to Type II error, if a difference is noted, is more likely to be of greater effect.

The higher level of presence is an extremely important finding, as it is the main rationale for further exploration of this technology, as, through this mechanism, it offers the potential for higher engagement, learning and retention.

Limitations

An argument can be made for constructing a study whereby learners use one system followed by the other, in a randomised way. The responses can then be measured using instruments to directly compare both systems. This method was not chosen as it was hypothesised that the novelty of the virtual reality system may incorrectly increase the difference between groups in favour of the immersive delivery method. In any case, the results have shown a statistically significant difference, therefore, this argument could only be used to describe a potentially larger effect than we have demonstrated.

3.7 Conclusions

This study showed that the simulator developed was usable and had face validity, in both delivery modes. Our results showed that there was a higher cybersickness in the immersive group, but this was not statistically significant and did not result in participant withdrawal. The most significant finding is the fact that using an immersive delivery method led to a greater degree of presence, which justifies the rationale in exploring the immersive delivery method further.

The next chapter presents work on creating an objective scoring system for the simulator and describes the creation of further modules to allow for further studies to be performed.

Chapter 5 – Assessing the contrast validity of the simulator and scoring system

5.1 Introduction

The aim of this chapter is to assess the ability of the simulator and scoring system to differentiate different levels of trainee (novice group, intermediate group, experienced group). The expectation is that the performance score will be higher for more experienced participants. As described in Chapter 1, the development of illness scripts from novice to expert is characterised by a development in the way information is transcribed by the individual. As such, SQ's have been used a triangulation tool and will serve to validate the accuracy of the performance score. I.e. the most senior trainees should have more well developed illness scripts and SQ's, and also higher performance score, than the junior trainee/medical student. This is in line with established literature and should serve to cross-validate our approach.

Comparing the usability and enjoyment of the simulator will build on work from previous chapters, suggesting that there should be no difference between the groups in terms of enjoyment or usability. Finally, self-confidence, as measured on a questionnaire, will be measured across all group, with the expectation that more experienced participants will have higher self-confidence ratings.

Hypothesis

Given the increased clinical experience and exposure to post-operative surgical patient management, we would expect the core surgical trainees to have the highest score, followed by the foundation trainees, with the medical students scoring the lowest.

5.2 Aims

The main aim of this study is to assess the contrast validity of the simulator

Specific objective:

1. To examine whether trainees score differently when using the simulator based on their level of experience.
2. To assess the ability of the semantic qualifiers to differentiate participants based on level of experience
3. To compare the enjoyment of the simulator by different participants based on level of experience
4. To compare the usability of the simulator by different participants based on level of experience
5. To compare the self-confidence of participants based on level of experience

5.3 Methods

5.3.1 Study design

This was a prospective cohort study assessing simulator performance on three different groups of trainees: medical students, foundation year doctors and final year core surgical trainees.

5.3.2 Setting

This study took place in the Patterson Centre at Imperial College London. Ethics was gained from ICREC and MEEC.

5.3.3 Randomisation

Participants were not randomised, all underwent the same conditions. All participants were given their consent form / questionnaire sheets with a unique random number, which was linked with the video recording of their attempt. The aim of this was to reduce observer bias by blinding the assessor to the training level of the subject.

5.3.4 Participants

Final year medical students (least experienced) - 16

Foundation year 2 trainees - 21

Core Surgical Trainees Year 2 (most experienced) - 19

5.3.5 Measurement instruments

In order to allow trainees to be scored, the entire case was recorded from start to finish for each individual, which allowed for all interactions to be assessed. The recorded videos were labelled and matched to the random number assigned on the consent form, as described above. The videos were reviewed and each interaction was transcribed into Excel, subsequently each interaction was marked against the scoring scheme created in Chapter 4. The breakdown of the scores were transcribed into Excel and imported into Stata for analysis.

System Usability Scale – as described and used in Chapter 3.

Presence Questionnaire – as described and used in Chapter 3

Face validity– as described and used in Chapter 3

Enjoyment questionnaire– as described and used in Chapter 3

Self confidence rating scale

5.3.6 Equipment

Google Daydream headset and controlled with Google Pixel 2 headset

5.3.7 Study process

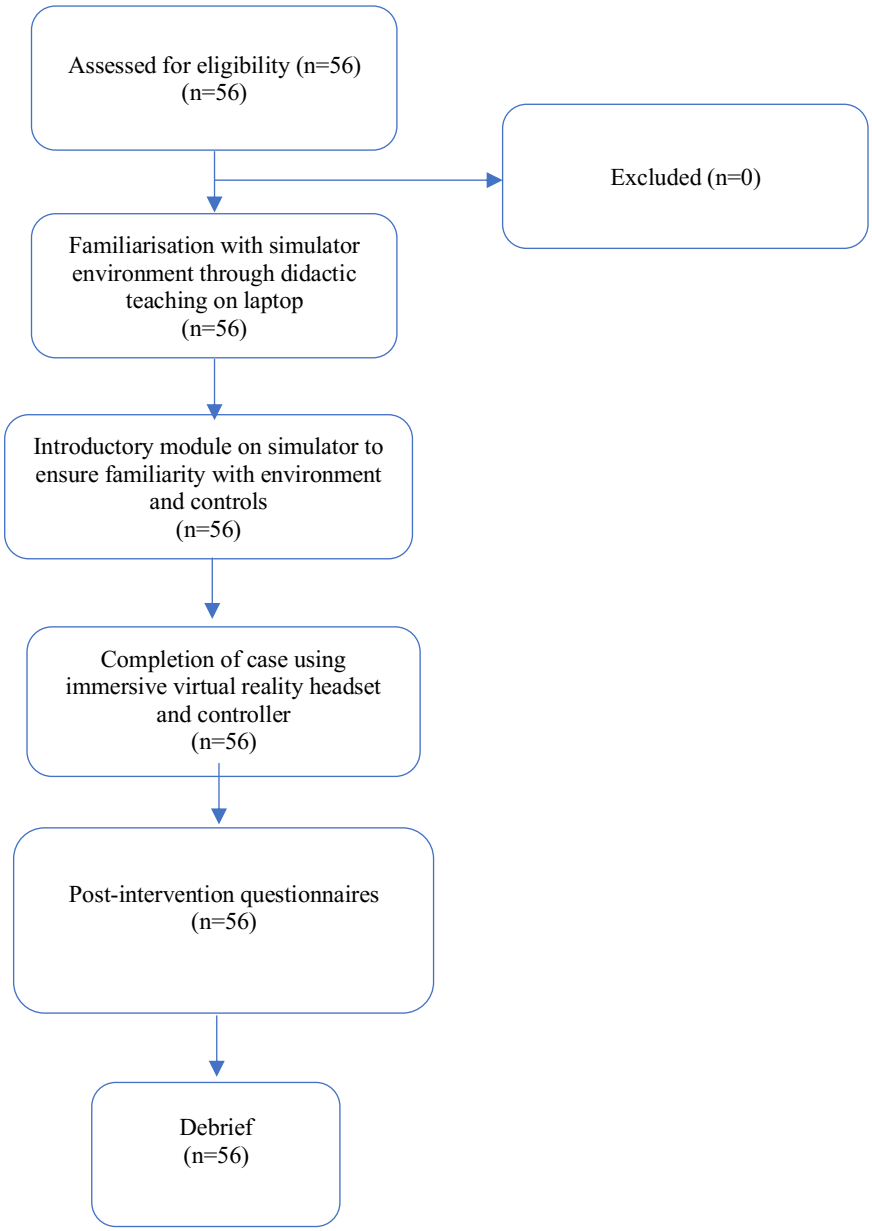


Figure 26: Study process

5.3.8 Statistical analysis

All data from the written questionnaires were transcribed to Excel and imported into Stata v13.0. Analysis of one way variance (ANOVA) test with Bonferroni correction was used, as this test accepts slight departures from normal variance.

5.4 Results

5.4.1 Performance scores

Table 8: Mean performance scores and standard deviation (SD)

Level	Mean performance score	SD	Frequency	M:F	Age
Core Trainee 2	57.3	10.1	19	14:5	28.8 (27-32)
Foundation Year 2	51.4	9.8	21	9:12	25.7 (24-29)
Medical student	46.4	9.7	16	7:9	23.7 (23-24)

Table 9: One-way analysis of variance of performance core by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	1030.0	515.0	5.29	0.0081
Within groups	53	5163.3	97.4		
Total	55	6193.2	112.6		

Analysis of variance showed that the effect of experience on performance score was significant, F (2, 53) = 5.29, p = 0.0081.

Post hoc analysis of performance scores between groups

Post-hoc analysis showed that there was a statistically significant difference in the performance score between the experienced group (57.3%) versus the novice group (46.4) ($p<0.006$). There was a difference between CT2 and FY2, with scores of 57.3% vs 51.4% though this did not reach statistical significance ($p=0.4$). Similarly, there was a difference in score between FY2 and MS, with scores of 51.6% vs 46.4% which did not reach statistical significance.

Sub-group analysis did not show any difference within groups based on previous VR experience (CT2 group, previous VR exposure vs none, 58.7 (n=8) vs 56.2 (n=11) ($p=0.32$), FY2 analysis not performed as only 1 participant had previous VR exposure, Medical Student group 45.7 (n=8) vs 47.1 (n=8) ($p=0.39$).

5.4.2 Semantic qualifiers

Table 10: Mean semantic qualifier scores, standard deviation (SD), and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	3.7	0.5	19
Foundation Year 2	3.4	0.7	20

Medical student	1.5	0.5	10

Six medical student results and one FY2 result were excluded as they did not fully complete the summarisation task, making it impossible to score accurately.

Table 11: One-way analysis of variance of semantic qualifier score by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	33.3	16.7	44.7	<0.05
Within groups	46	17.2	0.4		
Total	48	50.5	1.05		

Analysis of variance showed that the effect of experience on semantic qualifier score was significant, $F(2, 46) = 44.7$, $p < 0.05$.

Post hoc analysis of performance scores between groups

The above results showed that there was a statistically significant difference between both the CT2 and the medical student group ($p < 0.05$), and the FY2 group and the medical student group ($p < 0.05$). Similar to the performance scores, the CT2 groups did score higher than the FY2 group, but this difference was not statistically significant ($p = 0.28$).

5.4.3 Enjoyment of simulator use

Table 12: Mean enjoyment scores, standard deviation (SD), and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	16.3	1.9	19
Foundation Year 2	16.1	4.3	21
Medical student	16.8	2.9	16

Table 13: One-way analysis of variance of enjoyment scores by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	4.8	2.4	0.22	0.8
Within groups	53	568.4	10.7		
Total	55	573.1	10.4		

Analysis of variance showed that the level of experience did not affect the enjoyment of the simulator, $F(2, 53) = 0.22$, $p = 0.8$.

5.4.4 Usability of the simulator

Table 14: Mean system usability scores, standard deviation, and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	74.3	12.5	19

Foundation Year 2	80.5	12.3	21
Medical student	75.8	8.6	16

Table 15: One-way analysis of variance of system usability scores by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	411.6	205.8	1.57	0.2
Within groups	53	6939.7	131.0		
Total	55	7351.3	133.7		

Analysis of variance showed that the effect of experience on system usability scores was not significant, $F(2, 53) = 1.57$, $p = 0.2$.

5.4.5 Self-confidence ratings

Table 16: Mean self-confidence scores, standard deviation (SD), and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	3.70	0.43	19
Foundation Year 2	3.18	1.21	21
Medical student	3.25	0.59	16

Table 17: One-way analysis of variance of self-confidence scores by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	3.07	1.53	2.14	0.13
Within groups	53	38.0	0.72		

Total	55	41.0	0.75

Analysis of variance showed that the effect of experience on system usability scores was not significant, F (2, 53) = 2.14, p = 0.13.

Incorrect responses

Table 18: Mean percentage of incorrect responses, standard deviation, and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	1.54	1.95	19
Foundation Year 2	2.38	2.65	21
Medical student	4.97	4.98	16

Table 19: One-way analysis of variance of mean percentage of incorrect responses by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	109.1	54.5	4.97	0.0105
Within groups	53	581.9	12.0		
Total	55	691.0	12.6		

Analysis of variance showed that the effect of experience on the number of incorrect responses was significant, F (2, 53) = 4.97, p = 0.0105. The data showed that the CT2 error rate is less than both the FY2 and the MS, with errors of 1.5%, 2.4% and 5% respectively. The

error rate between the CT2 and MS group was statistically significant (p=0.011), with the error rate between CT2 and FY2, and between FY2 and MS being insignificant (p >0.05).

Comparison of correct responses

Table 20: Mean percentage of correct responses, standard deviation (SD), and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	58.8	9.36	19
Foundation Year 2	53.8	8.34	21
Medical student	51.4	7.21	16

Table 21: One-way analysis of variance of mean percentage of correct responses by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	510.7	255.3	3.6	0.034
Within groups	53	3748.1	70.7		
Total	55	4258.8	77.4		

Again the data shows the CT2 cohort had a higher percentage of correct responses versus FY2 and MS group (58.8% versus 53.8% versus 51.4% respectively), though this was only statistically significant between the CT2 and MS group (p=0.037).

Comparison of total responses

Table 22: Mean number of total responses, standard deviation (SD), and frequency by group

Level	Mean	SD	Frequency
Core Trainee 2	39.3	3.18	19
Foundation Year 2	39.5	3.19	21
Medical student	40.1	4.54	16

Table 23: One-way analysis of variance of total responses by level of experience

Source	df	SS	Mean Square	F	P
Between groups	2	6.7	3.34	0.26	0.78
Within groups	53	694.7	13.11		
Total	55	701.4	12.75		

There was no statistically significant difference in the total number of responses between the groups.

Figure 27: Showing use of immersive simulator during this study

5.5 Discussion

Performance scores

The main aim of this study was to establish if the simulator and scoring system could differentiate between more experienced and novices using our developed scoring system. In this first aim, it was successful, with the difference between the most experienced group (CT2) and the novice group (MS) being statistically significant. This is to be expected, as the literature suggests that clinical reasoning is a skill that continually improves through career of a junior doctor.

Whilst we did demonstrate a difference between CT2 and FY2, this did not reach statistical significance. Two possible conclusions can be drawn from this. Firstly, there is a difference between the groups, but the numbers required would be greater to reach statistical significance. Interestingly, the difference between the FY2's and medical students was not statistically different, which could suggest again that a large sample size is required, or that it takes four years of post-graduate training to develop the clinical reasoning skills exhibited by the Core Surgical Trainees. The results can be used to conduct a formal power calculation for further studies. Whilst this presents and interesting questions, the aim of this thesis is not to define the rate of acquisition of clinical reasoning skills in post-graduates, and therefore this work will not be investigated further.

The second conclusion that could be drawn is that there is significant improvement in clinical reasoning skills during the two foundation years of training, during which time trainees are actively carrying out assessments of patients with post-op pyrexia, resulting in no significant difference at this stage. As the simulation does not assess specific surgical knowledge, one

could surmise that the specific clinical reasoning skills involved in the assessment and management of a post-operative pyrexia surgical patient have been developed at an earlier stage.

The above finding also correlates well with the self-confidence expressed by difference groups. Again, there was a significant difference between the most senior and most junior, but not between the FY2 scores and the other groups.

Semantic qualifiers

This metric was used as a method of triangulation, as such, it correlates with the finding for hypothesis 1. In the same way as the Hypothesis 1 results, the SQ's between the CT2 and MS groups were found to be statistically significant. It must be noted that a significant number, around 6, of the MS summaries were excluded from analysis (and 1 FY2), all due to incomplete summarisation. Whilst it is possible that this was due to limited time availability or an incomplete understanding of the instructions, the author believes it is more likely that the MS group lacked the ability to appropriately and completely summarise the case. This would be in keeping with the literature, whereby, a novice would have completed the case mainly using System 2 thinking, thereby breaching cognitive capacity load early on, which would have reduced their ability to recall specific details. In addition, the ability of a novice to 'chunk' information is inferior, as presented in the opening chapter in relation to chess players. For these reasons, it is likely that the incompletion rates are a reflection of the inability to carry out the exercise. Despite this loss of numbers, the difference between the CT2 group and MS was statistically significant. Again, the difference between the CT2 and FY2 group was not

significant. This is again in keeping with H1 results, however, the difference here appears to be smaller. Contrary to H1, the FY2 SQ scores were statistically significantly higher than the MS group. It is likely that this reflects the early acquisition of semantic qualifiers during clinical practice. This highlights one of the drawbacks of using the SQ as a marker of ability in clinical reasoning, i.e. they are well developed *independent* of the contextual knowledge of the case. I.e. a medical registrar would most likely summarise the case in a similar way to a surgical registrar, thereby scoring the same on the SQ, but with a potentially different understanding of the case and different management strategy.

Usability and enjoyment The results build on the data gathered in Chapter 3 and demonstrate no difference in SUS and enjoyment between the groups.

Self-confidence

This was a very interesting finding. Whilst the results did in fact show that the self-confidence rating of the CT2 was higher than the MS, it was not statistically significant, and almost incredibly, the rating of the MS was higher than the FY2, despite them having less experience and scoring more poorly on the assessment. The author believes this highlights the inadequacies of self-confidence as a marker. Whilst it may perhaps be useful in monitoring the utility of an educational intervention, if done in a pre- / post- test manner, it does not in itself correlate to performance. This again is in keeping with the literature and is highlighted by both the Dunning-Kruger effect and the Johari Window concept (137)(138).

Score results

When looking at the breakdown of responses, the basis of the scoring system was to control for the shotgun approach. Interestingly, our data shows no statistically significant difference in the total number of responses between the groups, 39.3, 39.5, 40.1 (CT2, FY2, MS respectively). This was an unexpected result, based on the literature it was assumed that more inexperienced individuals would have more total responses as they would less sure about each one.

The data show that rather than a quantitative difference between groups, with the least experience group generating the highest number of queries, the distribution of responses between was qualitatively different. Our findings showed that the more experienced trainees were asking the same number of queries as the least experienced, but were simply choosing more correct answers, less irrelevant and less incorrect answers. The specific reasoning for this is unclear, as the number seemed stable across all groups and does not reflect the total number of enquiries possible in the system, i.e. a user could, if they wanted, generate far more responses than seen in the results. One explanation is that the final year medical students may have sufficiently developed clinical reasoning skills to know which options they obviously shouldn't choose, i.e. an appreciation that a CT head is not appropriate in a post-op surgical patient with pyrexia. One could test this hypothesis by assessing first year medical students to see if the total enquiries generated demonstrate a 'shotgun approach', as described in the literature. This finding does suggest a future scoring rubric does not need to take into the total number of responses, it could therefore be more specific and incorporate another important variable such as 'irrelevant' responses.

Limitations

It must be noted that the consensus gained by the expert panel (as discussed in the previous chapter), whilst unequivocal, was conducted by doing a walk-through of the case on a 2D computer screen with the primary researcher. It is possible that due to the familiarity with the system, the expert panel were exposed to all options available, whilst some learners may have 'missed' or not fully appreciated the full gamut of options available, due to the novelty of the technology. The findings from our initial System Usability Test suggest this would not have had a significant impact.

5.6 Conclusion

This study demonstrated that the simulator was able to differentiate between levels of experience based on the performance score. It also showed that more experienced trainees have more structured and refined semantic qualifiers, which is in keeping with the literature. As expected there was no difference in usability or enjoyment based on experience. This study was carried out to continue the validation work for this simulator.

Chapter 6 – Comparing the effectiveness of the simulator against traditional teaching methods

6.1 Introduction

In this chapter, we compare the effectiveness of immersive virtual reality against case-based discussion for teaching clinical reasoning. As discussed in Chapter 1, case-based discussions are one of the ways to develop clinical reasoning skills. Interactive case-based discussions are currently used to teach clinical cases for first year Core Surgical Trainees in London, as a means of introducing a clinical component to the surgical anatomy teaching.

6.2 Assessment of clinical reasoning

A script concordance test (SCT) is a written test that specifically assesses an integral aspect of clinical reasoning, i.e. the ability of the trainee to interpret information under the real-life conditions of uncertainty (139). The uncertainty in the case is integral and is created by developing cases where there is not necessarily one specifically correct answer for any given question. Following completion, trainees responses are scored against a reference panel of experts, with the score reflecting concordance (140). The theory is built on script theory, which is in line with the basis of this thesis, and therefore gives weight to the way in which knowledge is stored and retrieved (139). As a physician moves through the clinical reasoning process from differential diagnosis generation to management, they reference stored 'scripts'

and make judgements based on each new piece of information obtained (140). The SCT can simply be seen as a tool that allows the testing of each of these steps in reference to a panel of experts, thereby giving a measure of the trainees clinical reasoning ability. Questions should have neither low variability, as these are less discriminatory, or high variability which are not useful due to spread of the answers (140). It is recommended that in practice, lower variability questions, i.e. one's with answers to well-defined problems should be used (140).

The questions can be written to assess the diagnostic or management aspects of a case, creating uncertainty therefore requires careful consideration in construction to ensure efficacy.

As such, participants in both groups will have similar scores for the clinical reasoning problem questions as both case-based teaching and immersive VR allow for the generation of appropriate differentials and straightforward investigations. In addition, participants in both groups will have lower scores for the 'new' case for both the CRP and the SCT – demonstrating the context specificity of learning gains.

6.4 Hypothesis

Hypothesis **-** Participants in the treatment group will have higher scores on the script concordance test – the VR presents users with more invalid options, therefore requiring a more nuanced judgement of the relative importance of each.

6.3 Aim

The aim of this chapter was to compare the learning effects in clinical reasoning of the currently used case-based discussions against our virtual reality simulator

Objectives

1. Compare clinical reasoning score gains pre- and post-intervention with both VR simulator and traditional teaching
2. To examine the specificity of learning by assessing clinical reasoning score gain on a 'new' case

6.5 Methods

6.5.1 Study design

This was a prospective randomised controlled trial comparing learning gains between case-based discussion teaching and an immersive virtual reality teaching.

6.5.2 Setting

This study took place at the Imperial College Anatomy lab in Charing Cross Hospital. The study was conducted during one of three sessions that participants rotated through, the other two sessions were anatomy teaching through prosection. The time for each session was 50 minutes with each group size varying between 4 and 6. All immersive virtual reality sessions were delivered by Mr. Anoop Prasad, all interactive case-based discussions were delivered by Professor Nigel Standfield.

Ethics

This study was approved for ethics from ICREC.

6.5.3 Randomisation

Following consent, participants were randomly assigned numbers, even numbers were allocated into iVR with even numbers allocated into interactive CBD.

6.5.4 Subjects/Recruitment

First year core surgical trainees were chosen for this study, in total 53 trainees took part in this study. All trainees were consented for participation in this study. No trainees declined to participate, none met the exclusion criteria (non-CT1 trainee, previous adverse reaction to VR).

6.5.5 Assessment tool

6.5.5.1 Creation of a script concordance test

The guidelines by Fournier et al. were followed for test construction (140), the relevant points have been presented below.

1. Determine the purpose of the test – in this case, the test is to be used to assess learning gains after an educational intervention (control or simulator). It should therefore be composed of medium-difficulty questions which are able to discriminate the cohort being assessed (140).

2. Content validity – how representative of everyday practice are the questions / cases used. The cases were chosen by a panel of senior General Surgical Consultants to reflect common conditions seen on general surgical wards.

3. The role of uncertainty – this exists at two levels. Firstly by design it is introduced into the case. Secondly it is based on individual expert interpretation. The degree of uncertainty can be assessed from the overall panel scores, from -2 to +2 scores, with higher scores reflecting higher levels of certainty (140). Fournier et al. suggest that whilst high consensus answers have low variability and therefore lower discriminatory power, they are more useful in practice than high variability questions. In adherence to this recommendation, any answers with very low agreement were excluded.

4. Case content – the cases should be short enough that an expert is unable to give one clear answer, as either there is insufficient information, multiple options would be reasonable, or no consensus exists in the literature (140). This was assessed during the pilot stage of SCT creation.

5. Scoring – this is done by comparing the examinees response against the consensus panel using an aggregate scoring method (141). This involves having the number of respondents to the modal answer as the denominator, against which each test item is scored. For example, if five members of an expert panel consisting of ten member chose -1, with three choosing -2 and two choosing 0, the scoring for each would be 5/5 for -1, 3/5 for -2 and 2/5 for 0, with +1 a +2 receiving 0 marks.

6.5.5.2 Writing the questions

A group of three surgeons (two MRCS qualified and one FRCS nationally recognised trainer) wrote a series of script concordance test questions in line with the guidelines described

above. The cases were based on both the clinical scenarios developed for the simulator and other common surgical presentations, in total 12 cases were created.

6.5.5.3 Validation

Four cases were selected from this and tested as per the script concordance test item quality grid adapted from Caire (please see Appendix X for full checklist).

6.5.5.4 Scoring

Having demonstrated a high quality for all questions, the scoring was done with ten general surgical consultants to establish the marking scheme for subsequent studies.

Having two cases, one covering the topic covered by both the control and treatment intervention, with another different clinical case will demonstrate the context specificity of the knowledge. The pre-test questionnaire will demonstrate if there is a difference in knowledge between the two groups, which is not expected as the trainees are in the same training year. By assessing learners using both the script concordance test and the clinical reasoning problem, it allows us to further define the part of clinical reasoning that is being taught/has improved.

6.5.5.5 Sample Script concordance test

An example of a developed script concordance test is shown below:

Please complete the following questions, indicating if the statement in the third column is:

-2 much less likely

-1 less likely

0 – has no effect

+1 – more likely

+2 – much more likely

60 year old man 3 days post emergency open right hemi-colectomy spikes a temperature of 38.5C, high NG output

If you were thinking of:	And then you find	The diagnosis becomes				
Hospital acquired pneumonia	The patient doesn't report a cough	-2	-1	0	+1	+2
Urosepsis	He has a catheter in situ draining clear urine	-2	-1	0	+1	+2
Anastomotic leak	He has not opened his bowel or passed flatus	-2	-1	0	+1	+2

6.5.6 Protocol

Following consent and randomisation, all participants completed the demographics and pre-test questionnaires. This began with a demonstration of the case on a large screen projector, showing all the menu options till candidates felt comfortable with the layout.

6.5.6.1 Immersive virtual reality group

Participants in the treatment group were oriented to the immersive simulator by the lead researcher. They were given an introductory lecture by the researcher to familiarise them

with the simulator and its use, this includes completion of an introductory module using the simulator to ensure they are comfortable with the equipment. Following this all participants went through a post-operative pyrexia clinical scenario (post-op pneumonia case – presented in chapter 4). Finally the group takes part in a debriefing session whereby the case is discussed and rationale for each decision and management is reviewed.

6.5.6.2 Interactive case-based discussion group

Participants were given an introductory lecture by the researcher to familiarise them with the interactive case-based discussion and its use. Following this the participants were taken through the case by a senior clinician. The interactive case-based discussion was created by the researcher and is part of the current curriculum of teaching for post-graduate trainees.

All data given, including history, examination findings and investigations were identical to that of the immersive virtual reality case.

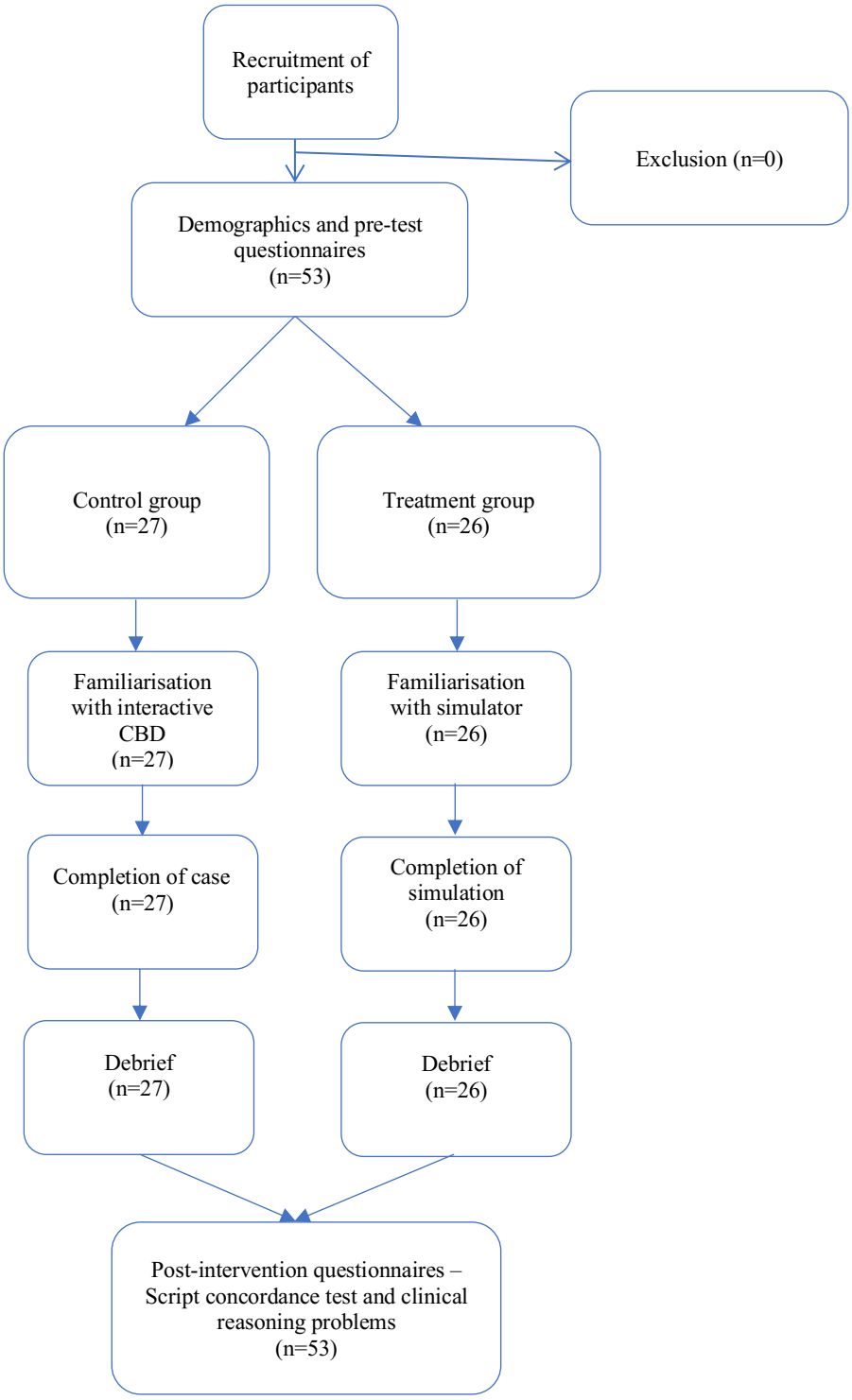


Figure 28: Study process

6.6 Results

Table 24: Participant demographics

Group	Male	Female	Mean age
Treatment	15	11	28.5
Control	16	9	28.3

Table 25: Grouped results for the control and treatment group showing the scores for the pre-questionnaire, Likert Enjoyment scale, clinical reasoning problems (CRP) 1 and 2, and script concordance tests (SCT) 1 and 2

	Control group	Treatment group	Significance
Pre-questionnaire	4.84	5.5	p=0.21
Likert	16.08	16.81	p=0.38
CRP1	4.24	4.46	p=0.42
CRP2	2.85	3.03	p=0.33
SCT1	1.64	2.35	p=0.045
SCT2	2	2.27	p=0.52

The only statistically significant difference was in the Script Concordance Test 1, which was based on the case that was taught to both groups. In this, the Treatment group scored significantly higher than the control group (2.35 vs 1.64, p = 0.045). In all the other comparisons, including pre-test scores, the control group and the treatment group did not have statistically significant differences.

Table 26: Grouped results for the control and treatment group showing the scores for the enjoyment, usefulness for interpreting tests and improving decision-making and the trainees motivation to use a library of cases if given access

Likert questionnaire	Control group	Treatment group	df	t	significance
Enjoyed	3.96	4.308	41.6	1.704	p=0.096
Use library	3.8	4.039	45.08	0.987	p=0.33
Useful for interpreting simple tests	4.08	4.153	43.56	0.322	p=0.75
Useful to improve clinical decision-making	4.24	4.31	38.7	0.312	p=0.76

There was no statistically significant difference between the control and treatment groups with regards to enjoyment or perceived usefulness.

Figure 29: Pictures showing the use of the immersive virtual reality simulator and the introduction to the simulator

6.7 Discussion

Hypothesis 1 **Participants in the treatment group had higher scores on the script concordance test**. This is an important finding as the main aim of this thesis has been to develop a training tool that can improve clinical reasoning. The results show that the iVR simulation better teaches clinical reasoning versus the traditional case-based discussion. The specific mechanism by which this occurs is unclear, but is most likely due to a number of reasons, which are consistent with the literature.

Firstly, and as shown in Chapter 3, the iVR system creates increased presence. This was true against a VR system in 2D and we would therefore expect an even larger difference in presence for the iVR system versus an interactive CBD. Increased presence increased both

recall and situational awareness within a simulation. In addition, the increased realism of the iVR system allows for a closer alignment of real-world and in-simulation behaviour, allowing for the learner to engage in near-transfer rather than far-transfer (which is the case for the iCBD). One specific mechanism that highlights this is the ability for the simulator to display almost all the diagnostic and therapeutic interventions available in the real world. In doing so, it forces the learner to differentiate between what is most relevant and what is less relevant, which is the basis of clinical reasoning and also the fundamental principle underpinning of the SCT. Unfortunately, the available display real-estate in the interactive CBD simply limits the number and complexity of options that are viewable at any one time. Inevitably this leads to a more structured or scaffolded approach being taken for the interactive CBD cases, which mitigates the gains of the iVR system.

In addition, the iVR system engages the learner more than the interactive case-based discussion format. I.e. in the virtual reality simulator, each learner is forced into engaging with the clinical scenario, by selecting available options which reinforces the learning of both the correct and incorrect options. This process in the clinical reasoning process allows the learner to acquire an understanding of the nuances between each test or diagnosis. As each learner is more involved in the case, this most likely increases the value of the facilitator-led discussion, as all learners will be able to reflect on the correct choice presented versus the choice they had made. This is in contrast to an interactive CBD, where fewer learners can be actively engaged in answering questions at any given time.

The SCT addresses five out of the six steps of clinical reasoning, as described Kassirer et al. and is therefore a more accurate marker of clinical reasoning than the clinical reasoning problem presented below. It should be stated that a key aspect not assessed by the SCT is the

ability to generate a list of differentials, though, through second order questioning, require the learner to have chosen the correct most likely diagnosis in order to answer all subsequent questions accurately.

Hypothesis 2 **Participants in both groups will have similar scores for the clinical reasoning problem questions** – the clinical reasoning problems (CRPs) were used in this study to help identify which aspects of clinical reasoning were being taught superiorly by the iVR system versus the interactive CBD system. In line with the literature, the participants in both groups had similar scores for the clinical reasoning problem questions. This is simply because the CRPs are designed to assess the ability to generate a list of differential diagnoses and appropriate tests for a given presentation, which are steps 1 and 4 in Kassirer's description of the clinical reasoning process (presented in Chapter 1). This is a key part of clinical reasoning that is taught by both the iVR group and the interactive CBD group, therefore, this finding is in keeping with this.

Hypothesis 3 **Participants in both groups had lower scores for the 'new' case for both the CRP and the SCT**. This is an important point to validate that has been shown in the literature. That is the fact that clinical reasoning is not a generic skill that can be learnt, but a context specific skills whereby exposure to a case predicts success in future exposures to that case, or at the very least, a very similar case. It does not confer a learning gain if the learner is faced with a completely different problem (48). This justifies the case for exposing a learner to a large library of cases in order to develop clinical reasoning in different contexts.

Limitations

A key limitation in the interpretation of this data is that two different instructors of two different levels, a Professor of Surgery (for interactive CBD) and a post-MRCS registrar (for iVR), were used to facilitate the teaching. This was a known limitation that was required due to the time constraints for conducting the study as part of the Deanery teaching programme. It is likely that the Professor of Surgery, with greater teaching experience, would be able to deliver a higher quality teaching session. The results did not show this, and if true, suggest that the learning gains seen in the iVR group could possibly be greater than has been demonstrated. Another limitation is that students may have been more intimidated by teaching delivered by a Professor of Surgery, which may have led to a reduced willingness to participate, and therefore reduced learning gains. Looking at the data on enjoyment, whilst the control group (interactive CBD group) did have a slightly lower mean score of 3.96 versus 4.31, this was not statistically significant. As described above and shown in Chapter 3, the iVR system likely lead to a greater level of presence in the content, and it may be that the quality of the discussion with the facilitator would have been greater, as learners would be more engaged in the case and therefore would have picked up all the finer details of the case, leading to a richer debate with the facilitator. This is not something that was measured and is an inherent difficulty when considering two different teaching modalities, as the behaviours around each one also differ, and are in fact a central part of the modality. A potential area for future study would be to investigate qualitatively the difference in the discussion around the cases between the two modalities.

6.8 Conclusion

This study has shown that the iVR simulator was more effective at teaching clinical reasoning than traditional case-based discussion. This was reflected in the higher post-test scores on

the script concordance test. The study also shows the differential generation phase of clinical reasoning, which was assessed through the clinical reasoning problems, was equivalent across both interventions, which was in line with initial expectations. Finally, the study showed, in line with current literature, the context specificity of the learning gains, whereby gains in one clinical case did not translate into gains for a completely different clinical case.

Chapter 7 - Discussion

7.1 Discussion of thesis by objectives

The work presented in this thesis started as an attempt to create an immersive virtual reality simulator using 360 degree video to teach and assess clinical reasoning. Specifically, we wanted to collect data in order to address the objectives below:

1. **To demonstrate feasibility of construction of an immersive virtual reality clinical scenario using 360-degree video to teach clinical reasoning**
2. **To assess the usability of a pilot scenario against a range of potential future users**
3. **To assess if increased immersion results in increased presence for when using virtual reality in a clinical scenario – this will assess two different delivery modes, a computer screen versus a head-mounted display**

First, the two initial studies aimed at creating the simulator and subsequently assessing it through testing its usability, face validity, presence and cybersickness. The first study successfully demonstrating that creating an immersive virtual reality simulator using 360 degree video was feasible, which allowed for the next stage of testing. During the scope of Study 1,the story-boarding was carried out in a similar structure to existing virtual patients, and the design principles focussed on avoiding simulator sickness, and establishing familiarity. The second study built upon the completion of Study 1, and the main focus was to successfully demonstrate that the simulator was usable, as measured through the System Usability scores, as well as that it avoided simulator sickness. With regards to usability, the results were

different to the hypotheses. The assumption before the study was that the usability of the system in the 'immersive head-mounted display' arm, would have lower scores, as this was an inherently more complex set-up, where the user would be interacting with a virtual world using a headset and controller, with which they were unfamiliar. In contrast the set-up on the laptop, had a more familiar set-up, with a mouse and keyboard. This result could potentially be or could be a reflection of the appropriate application of design principles, or, less likely, could be described by the age or plasticity for learning new technology of the participants. This is less likely as the participants ranged from medical student to senior registrar, with an age range of 19-50. It would perhaps be interesting to investigating if there was a perceived difference in usability between trainees and consultants.

The results also showed that there was no significant difference in cybersickness scores between the two groups, which was a key finding, given the reported rates in the literature of 22-80% (135). This again is likely a reflection of both the technology, the studies reporting these rates were using older technologies, and the application of the virtual reality design principles to reduce motion sickness.

The results also showed that both methods of delivery (head-mounted display and laptop screen) were equal in face validity. This was against our initial hypothesis where we believed a more immersive environment would appear more realistic of a clinic encounter, and therefore increase the face validity. It is likely that the study design, in which each participant only used one delivery method, lead to no difference being noted, or a Type 1 error, as they would have relativised their scores to the simulator being used. Whilst there is no literature to show this effect, it is likely that the face validity score would have been higher for the more immersive technology, if the study had asked the participants to go through both delivery methods for the simulator. Due to time constraints with participants, this was not possible.

Perhaps the most significant finding that guided the path of discovery for this thesis was the demonstration of greater levels of presence in the head-mounted display group. Despite the participants using the different delivery methods in isolation, and the reported high type II error rates of the most validated presence measurement instruments, the study was able to show a significant difference between the two delivery methods. This suggests that the difference in presence may be greater than demonstrated in this study. This result justified the continued investigation of this simulator in its immersive virtual reality form, because increased presence has been shown to increase recall and situational awareness in simulation (90,91). In addition increased presence results in participant behaviour that is more similar to the real world (92).

4. **Construct a series of clinical scenarios to allow for further studies, specifically scoring, and learning gains versus current standard practice;**
5. **Create a scoring system to allow for performance assessment whilst using the simulator;**

One of the main advantages of simulated patients is their use in assessment as they provide standardised exposures, and a high face and predictive validity that paper-based assessments are unable to provide (125). As such, the main objective of the third study was to create an objective scoring system, as well as three more modules to allow for further studies. Every interaction in a case is called an enquiry, and each enquiry was assigned three levels of appropriateness based on the case, correct, incorrect (potentially harmful) or irrelevant (not correct but also not harmful). As described in the literature, scoring rubrics need to be designed based on the objective of the assessment(127). As the objective of our simulator

was to assess final year medical students and post-graduate doctors, the scoring rubric chosen was specifically designed to both reward the correct answers, whilst also penalising for the 'shotgun' approach that would be expected with more junior learners. The three clinical scenarios were created with the same methodology as Chapter 2, and all covered post-operative sepsis, as it is a leading cause of multiple-organ dysfunction and in-hospital mortality(123). In addition, the characteristics of the assessment and management, specifically the diagnostic uncertainty and need for a range of differentials, made it an ideal topic for the teaching and assessment of clinical reasoning.

6. **To assess the ability of the simulator to differentiate between users of different clinical ability; and**

The fourth study aimed to assess the ability of the simulator and scoring system to differentiate levels of trainee based on experience. In line with our expectations, the more experienced trainees scored higher, but this difference was only significant between Core Surgical year 2 trainees (CT2) and final year medical students (MS). This is in line with the literature, which shows that clinical reasoning is a skill that continually improves through the career of a doctor. The reason for the non-significant difference between MS and foundation year 2 doctors (FY2), and between FY2 and CT2 doctors could be a type I error, and this study could be used to conduct formal power calculations in future studies. It could also represent the rate at which clinical reasoning skill is acquired, with a significant difference in clinical reasoning only present after several years of post-graduate training, once the trainee has formed well-defined illness scripts and exemplars for the cases in question (139). Because clinical reasoning is also context specific, rather than a generic skill (48), it is likely that this

timeline of acquisition will be different based on how frequently learners are exposed to the cases in everyday practice. As post-operative sepsis is a commonly encountered condition, this could represent the shorter time for the acquisition of clinical reasoning skills versus rarer conditions. The rate of acquisition of clinical reasoning skills for different conditions represents an interesting line of enquiry for future research. Specifically, because curricula are designed to highlight life-threatening and emergency cases, which may have a lower exposure, and therefore may warrant increased use of simulation to gain this experience and shorten the time to acquisition of clinical reasoning skill in that context. Semantic qualifiers (SQ's) were used as a triangulation method, as they have been shown to relate to diagnostic accuracy, and are therefore a proxy marker for expertise (59,60). In line with the literature, our study showed increasing complex semantic qualifier use by more senior trainees. The SQ scores showed the same trend as the performance scores, with the CT2 scoring the highest and the MS lowest, there was however a statistically significant difference between the CT2 and MS and the FY2 and MS groups. This again could be due to a type I error, or more likely, it represents the drawbacks of using SQ as a marker of ability in clinical reasoning. This is because SQ development does not necessarily align with contextual knowledge, i.e. it can be developed independently through experience (59,60). In either case, the similar trend of SQ's with the simulator performance score can be viewed as further evidence for construct validity as it seems to differentiate trainees across experience levels.

This study also assessed the self-confidence rating of trainees, with the CT2 group scoring higher than the MS. Though not statistically significant, and interestingly, the rating of the MS was higher than the FY2, despite them having less experience and scoring more poorly on the assessment. The author believes this highlights the inadequacies of self-confidence as a marker of competence. Whilst it may perhaps be useful in monitoring the utility of an

educational intervention, if done in a pre- / post- test manner, it does not in itself correlate to performance. This again is in keeping with the literature and is highlighted by both the Duning-Kruger effect and the Johari Window concept (137,138), where the medical student, with the least experience is at the highest level of confidence, versus the FY2, who having had more exposure to the clinical world, is more aware of how much they don't know, the so-called 'Valley of despair'(129).

7. **To assess the learning gains in clinical reasoning following use of the simulator versus the case-based teaching method currently employed in post-graduate training.**

The final study of the thesis assessed the simulator against current learning methods employed. In this study, it was used to teach clinical reasoning skills to post-graduate Core Surgical Trainees as part of their clinical anatomy teaching sessions. This study had two main components, firstly, the measurement instruments (script concordance test and clinical reasoning problems) were created using guidelines in the literature. The second component was delivering the teaching and assessing learning gains amongst the trainees versus the traditional teaching. The results showed that the participants in the treatment group had higher scores on the script concordance test. This is an important finding as the main aim of this thesis has been to develop a training tool that can improve clinical reasoning. The results show that the simulator is more effective at improving clinical reasoning than traditional methods. The likely explanation from the literature is that with simulator, there is increased presence, which allows for both increased recall and situational awareness. Secondly, the amount of display real-estate available allows for most of the real-world options to be

displayed, which increases the ability to differentiate what is relevant, and forces the trainee to think more critically about their chosen investigations and treatments. The results also showed that participants in both groups had lower scores for the 'new' case for both the CRP and the SCT – demonstrating the context specificity of learning gains. Furthermore, both groups had lower scores for the new case using both measurement instruments (CRP and SCT). This highlights that although the clinical reasoning had improved for the cases that were taught, there is a context specificity demonstrated in the learning gains. This is in line with the literature which shows that clinical reasoning is not a generic skill, but rather specific to the illness encountered(48). This finding further supports the need for the creation of large case libraries to allow for a broad development of clinical reasoning for common and uncommon conditions.

7.2 Implications and applications of the research

This thesis has demonstrated that it is feasible to create clinical scenarios using immersive virtual reality 360 degree videos. It has been demonstrated that these can effectively be used to develop clinical reasoning skills, and furthermore, it can be used to differentiate between trainees of different levels of experience.

7.3 Limitations of the thesis and future directions

One of the key assumptions is that the trainees behave as they would in the real world. This is of course an assumption and it is not possible to tell how trainees would have really reacted had this been a real case.

Three described effects are worth discussion. Firstly, hypervigilance in a simulation, whereby trainees are excessively concerned as they know that the simulation will throw up a situation that requires intervention. This effect perhaps means that trainees are more likely to pick up subtle clues as they are expecting the patient to have a recognisable pathology, which is not the case in real life (142,143). Secondly, the behaviour of learners can be cavalier as the risk of actual patient harm has been removed (142–144). Whilst these are known limitations, the conditions were the same for all groups involved, so should not have necessarily made a difference in this study. In addition, it has been argued that these two effects working together may cancel each other out (145).

Another key limitation that must be acknowledged is the expertise reversal effect, where the same instructional design is used on both novice and experienced learners (146). The novice learner, who requires a more structured representation of data, which may be less realistic but more accessible. For the experienced learner, a more structured, or simplified, case may cause them to become disengaged, reduce trust in the simulator, or not react as expected/planned to prompts clearly if they are not backed up by data points in the case. For example, a more senior trainee may suffer from cognitive dissonance if the oxygen saturations is low, but the patient does not look like a patient who is short of breath; or they may not believe that a patient is febrile, if they are not visibly sweating or don't look 'septic'. As we were not using professional actors in these cases, there is a possibility that there was inconsistency between the visual representation of the patient and data representation, which would have had a greater effect on more senior trainees. The end result of this is the possibility that the scores of the most senior trainees could have been reduced due to lack of engagement or trust in the simulator.

When assessing a new technology, it is important to discuss Vroom's work on the expectancy theory of motivation, which describes how an individual's belief about a task or technology, will influence their performance (147). The believed mechanism is a positively viewed technology will be accompanied by increasing levels of motivation and expectation, whilst a negatively viewed technology will be accompanied by the opposite, with reduced motivation and expectation. As immersive VR is presented in the mainstream media as a 'cool' technology, it is likely to be perceived in a positive light by the users, it is possible that the users were more engaged and motivated then they would have otherwise been. Whilst this would not be a problem normally, if this is the underlying mechanism behind increased motivation, it could also theoretically be reduced with increased exposure. This definitely warrants further investigation, in the form of studies where users have repeated exposure to the technology, for example, carrying out a course on the simulator over an extended period, which will allow this theory to be further examined. In Study 2, where the immersive VR was tested against the same platform on a normal screen, it is unlikely that this theory led to a difference in outcome. This is because the user feedback suggested that were equally impressed by the 2D as they were the immersive VR system. This may have had a greater effect if the study were constructed in such a way as to allow each participant to use both systems and compare, where it would seem reasonable to assume that they would expend more energy and show greater enthusiasm when using the more high-tech immersive virtual reality system.

7.4 Future directions

7.4.1 Measuring real-world effect

Whilst the thesis has demonstrated that the tools can be used to increase learning gains in clinical reasoning, it has not demonstrated that this will lead to improved patient outcomes, which is the ultimate aim of any medical training simulator. Future work in this direction could involve the use of similar cases prior to trainees starting on a new clinical rotation, and then assessing metrics such as never events, retrospectively assessing missed/false diagnoses or auditing compliance with national protocols.

7.4.2 Assessing learning decay

Another aspect that has not been investigated is the learning decay that occurs using this simulator. An interesting avenue of research would be to assess differences in the decay of the skill learned. The strong argument from the literature is that immersive virtual reality should lead to increased presence and engagement, which should make the interaction closer to real life, thereby, providing more stimulatory cues to increase the retention rates and reduce the decay of knowledge.

7.4.3 Reflection-in-action vs reflection-on-action

There are no data to support if a learner benefits most from reflecting on decisions at set periods during the simulation, or whether it is best served by a debrief at the end(148). Some preliminary data showed that junior learners preferred to be de-briefed at the end of each stage of the case, i.e. after history and examination, results review, or intervention, whereas more senior trainees preferred to be de-briefed at the end. Using the same level of trainee, a

study could be constructed, which looks at both the enjoyment and effectiveness (measured using clinical reasoning measurement tools) of de-briefing as decisions are made, so called 'reflection-in-action', versus de-briefing at the end of the session, or 'reflection-on-action.

7.4.4 Trialling in critical intensive learning periods

Learners transition through several stages in their career from medical student to independent practice as consultants or general practitioners. Traditionally, and understandably, these periods have shown to be perceived as daunting and the literature reports that it is linked with anxiety, depression and burnout (149). Studies have reported these negative emotions arise as a result of the sudden increase in level of responsibility in patient care (24,150–156).

Transitions have been described as critical intensive learning periods (CILPS) where the learner's performance is critically dependent on the working environment (24). Learning experiences are therefore more useful when learners gain practical 'know-how' rather than just theoretical knowledge. This lack of practical knowledge leads to negative emotions, as discussed earlier, and is compounded by inadequacy of inductions in introducing the learner to their day to day work (154)(156), rather, they tend to focus on organisational agendas.

Decision-making and prioritisation skills are important components in preparedness and transition (157), within this the role of realistic simulated ward exercises have been recommended in preparedness for practice (158,159).

Whilst there has been substantial interest in pre-transition training, with a growth in preparation courses, there is little to suggest that these achieve more than a short-term reduction in stress (24). Many of the studies assess confidence as a proxy marker for competence, despite there being no clear correlation. Furthermore, a study with longer follow-up has shown no advantage over the control group at two months (160).

7.4.5 Use for research into clinical reasoning

Perhaps one of the greatest uses of the simulator developed is its potential to develop and track decisions to allow for further research on clinical reasoning. If the user's gaze and every option chosen can be assessed, this may represent a significant area for further research as it may offer the potential to unlock decision-making in 'real-time' between experts and novices. Combining the user's gaze and think-out loud protocols could lead to further insights into how clinical reasoning occurs, allowing further validation or challenge to Croskerry's model.

The work constructed within the scope of this thesis paves the way for a potentially exciting new way to teach, assess and understand how clinical reasoning works, is developed and can be nurtured in a way to improve patient safety. It is also aligned to the current dynamics and changes in light of the pandemic, as it could be used as a tool for remote teaching and training.

7.4.6 Personal reflections

Working in a field that is a rapidly evolving as virtual reality presents certain challenges to the conduct of a PhD. A key piece of advice that I would give anyone embarking on research in this field is to establish a rapport with the technology partner/platform that you choose. This will allow you to have a clear idea of their future development line, with expected timelines.

In doing so, this allows you to plan and conduct your work and studies around these timelines. Another point to note is to look at how different industries and technologies have approached similar problems to the ones you are trying to solve. If you find guidelines are lacking in the specific application you are researching, look outside of just medicine or healthcare. In doing so, you will be able to link/defend your chosen path with the existing literature, even if it is from a different specialty area.

References

1. Medical Education in the United States and Canada | Enhanced Reader [Internet]. [cited 2021 Mar 24]. Available from: moz-extension://a5e42042-f396-184c-9c52-e13d77ddc06d/enhanced-reader.html?openApp&pdf=http%3A%2F%2Farchive.carnegiefoundation.org%2Fpublications%2Fpdfs%2Felibrary%2FCarnegie_Flexner_Report.pdf
2. Stone MJ. The wisdom of Sir William Osler. Am J Cardiol [Internet]. 1995 Feb 1 [cited 2019 Feb 6];75(4):269–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7832137

3. Norman G. Medical education: past, present and future. Perspect Med Educ [Internet]. 2012 Mar 7 [cited 2019 Mar 26];1(1):6–14. Available from: http://link.springer.com/10.1007/s40037-012-0002-7
4. Harden RM. Trends and the future of postgraduate medical education. Emerg Med J [Internet]. 2006 Oct [cited 2019 Mar 26];23(10):798–802. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16988312
5. Harden RM. Developments in outcome-based education. Med Teach [Internet]. 2002 Jan 3 [cited 2019 Mar 26];24(2):117–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12098427
6. HARDEN RM. AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. Med Teach [Internet]. 1999 Jan 3 [cited 2019 Mar 26];21(1):7–14. Available from:

http://www.tandfonline.com/doi/full/10.1080/01421599979969

7. David MF Ben, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. AMEE Medical Education Guide No. 24: Portfolios as a method of student assessment. Med Teach [Internet]. 2001 Jan 3 [cited 2019 Mar 26];23(6):535–51. Available from: http://www.tandfonline.com/doi/full/10.1080/01421590120090952
8. Hesketh EA, Bagnall G, Buckley EG, Friedman M, Goodall E, Harden RM, et al. A framework for developing excellence as a clinical educator. Med Educ [Internet]. 2001 Jun [cited 2019 Mar 26];35(6):555–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11380858
9. White C. GMC to introduce formal recognition for trainers, mentors, and educational supervisors. BMJ [Internet]. 2012 Aug 23 [cited 2019 Mar

26];345:e5741. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.e5741

10. Temple report backs consultant based service. BMJ [Internet]. 2010 Jun 11 [cited 2019 Mar 26];340:c3118. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.c3118
11. McIndoe AK. Modern anaesthesia training: is it good enough? Br J Anaesth [Internet]. 2012 Jul 1 [cited 2019 Mar 26];109(1):16–20. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0007091217329264
12. Goddard AF. The European Working Time Directive and the impact on training: the current evidence. Clin Med [Internet]. 2010 Aug [cited 2019 Mar 26];10(4):317–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20848999
13. Hartle A, Gibb S, Goddard A. Can doctors be trained in a 48 hour working week?

BMJ [Internet]. 2014 Dec 9 [cited 2019 Mar 26];349:g7323. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25498122

14. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise. Acad Med [Internet]. 1990 Oct [cited 2018 Oct 12];65(10):611–21. Available from: https://insights.ovid.com/crossref?an=00001888-199010000-00001
15. O'Dowd A. Funding boost aims to expand England's medical school places. BMJ [Internet]. 2021 Aug 11 [cited 2021 Nov 29];374:n1998. Available from: https://www.bmj.com/content/374/bmj.n1998
16. Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. Int J Med Educ. 2016;7:200–5.
17. S. T. "The Mozart of psychology" in Mind in Society: The development of higher

psychological processes by L.S. Vygotsky. Vol. 25. 1978. 51–57 p.

18. (PDF) Kozulin, A., Gindis, B., Ageyev, V., Miller, S. (Eds.) (2003). Vygotsky's educational theory in cultural context. Cambridge, England: Cambridge University Press. [Also French language edition, 2009] [Internet]. [cited 2021 Mar 24]. Available from: https://www.researchgate.net/publication/220009990_Kozulin_A_Gindis_B_Ageyev_V_Miller_S_Eds_2003_Vygotsky%27s_educational_theory_in_cultural_context_Cambridge_England_Cambridge_University_Press_Also_French_language_edition_2009
19. Yardley S, Teunissen PW, Dornan T. Experiential learning: AMEE Guide No. 63. Med Teach [Internet]. 2012 Feb [cited 2018 Jun 19];34(2):e102–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22289008

20. Nalliah S, Idris N. Applying the learning theories to medical education: A commentary.
21. Neumann Jr Maurice A Deane RK, Neumann Jr RK. Deane School of Law at Hofstra University Scholarly Commons at Hofstra Law Hofstra Law Faculty Scholarship Spring. Vol. 401, Rev. 2000.
22. Jorwekar GJ. Reflective practice as a method of learning in medical education: history and review of literature. Int J Res Med Sci. 2017;5(4):1188.
23. David A. Kolb. Experiential Learning: Experience as the Source of Learning and Development: Amazon.co.uk: Kolb, David A.: Books. 1st ed. Prentice Hall; 1983. 41 p.
24. Yardley S, Westerman M, Bartlett M, Walton JM, Smith J, Peile E. The do's, don't and don't knows of supporting transition to more independent practice. Perspect Med Educ [Internet]. 2018 Feb [cited 2018 Jun 19];7(1):8–22. Available from:

http://link.springer.com/10.1007/s40037-018-0403-3

25. Miller GE. The assessment of clinical skills/competence/performance [Internet]. Vol. 65, Academic Medicine. 1990 [cited 2020 Feb 20]. p. S63–7. Available from: http://journals.lww.com/00001888-199009000-00045

26. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. Int J Med Educ. 2019 Oct 25;10:191–2.

27. Ericsson KA, Krampe RT, Tesch-Romer C. The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychol Rev. 1993;100(3):363–406.

28. Anders Ericsson K, Harwell KW. Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: why the original definition matters and recommendations for future research. Front Psychol.

2019;10(OCT):2396.

29. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med [Internet]. 2004 Oct [cited 2018 Oct 25];79(10 Suppl):S70-81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15383395

30. Kevin R. Murphy. Psychological Testing [Internet]. Fifth. Pearson Education (US); 2000 [cited 2021 Mar 24]. Available from: https://www.worldofbooks.com/en-gb/books/kevin-r-murphy/psychological-testing/9780130273956?gclid=EAIaIQobChMI5OT0o6DJ7wIVu2DmCh1zfA-uEAQYBCABEgI3yvD_BwE

31. Aucar JA, Groch NR, Troxel SA, Eubanks SW. A review of surgical simulation with attention to validation methodology. Vol. 15, Surgical Laparoscopy, Endoscopy and Percutaneous Techniques. Lippincott Williams and Wilkins; 2005. p. 82–9.

32. Technical recommendations for psychological tests and diagnostic techniques. Psychol Bull. 1954 Mar;51(2 PART 2):1–40.
33. Paul Kline. The New Psychometrics: Science, Psychology and Measurement [Internet]. 1st ed. Routledge; 1999 [cited 2021 Mar 24]. Available from: https://www.routledge.com/The-New-Psychometrics-Science-Psychology-and-Measurement/Kline/p/book/9780415228213
34. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, et al. Consensus guidelines for validation of virtual reality surgical simulators [Internet]. Vol. 19, Surgical Endoscopy and Other Interventional Techniques. Surg Endosc; 2005 [cited 2021 Mar 24]. p. 1523–32. Available from: https://pubmed.ncbi.nlm.nih.gov/16252077/
35. Alsalamah A, Campo R, Tanos V, Grimbizis

G, Van Belle Y, Hood K, et al. Face and content validity of the virtual reality simulator 'ScanTrainer®.' Gynecol Surg [Internet]. 2017 Dec 1 [cited 2021 Mar 24];14(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28959176/

36. McDougall EM. Validation of Surgical Simulators. J Endourol [Internet]. 2007 Mar [cited 2019 Jan 21];21(3):244–7. Available from: http://www.liebertpub.com/doi/10.1089/end.2007.9985

37. McDougall EM, Corica FA, Boker JR, Sala LG, Stoliar G, Borin JF, et al. Construct Validity Testing of a Laparoscopic Surgical Simulator. J Am Coll Surg [Internet]. 2006 May [cited 2021 Mar 24];202(5):779–87. Available from: https://pubmed.ncbi.nlm.nih.gov/16648018/

38. Newman-Toker DE, Wang Z, Zhu Y, Nassery N, Saber Tehrani AS, Schaffer AC,

et al. Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: Toward a national incidence estimate using the “big Three.” Diagnosis [Internet]. 2021 Feb 1 [cited 2021 Nov 29];8(1):67–84. Available from: https://www.degruyter.com/document/doi/10.1515/dx-2019-0104/html

39. Zwaan L, De Bruijne M, Wagner C, Thijs A, Smits M, Van Der Wal G, et al. Patient Record Review of the Incidence, Consequences, and Causes of Diagnostic Adverse Events. Arch Intern Med [Internet]. 2010 Jun 28 [cited 2021 Nov 29];170(12):1015–21. Available from: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/416102
40. Kassirer JP, Wong JB, Kopelman RI. Learning clinical reasoning. Lippincott Williams & Wilkins Health; 2010. 329 p.
41. Nendaz M, Perrier A. Diagnostic errors and flaws in clinical reasoning:

mechanisms and prevention in practice. Swiss Med Wkly [Internet]. 2012 Oct 23 [cited 2018 Jun 8];142:w13706. Available from: http://doi.emh.ch/smw.2012.13706

42. Neale G, Woloshynowych M, Vincent C. Exploring the causes of adverse events in NHS hospital practice. J R Soc Med [Internet]. 2001 Jul [cited 2018 Jun 8];94(7):322–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11418700

43. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med [Internet]. 2005 Jul 11 [cited 2018 Jun 8];165(13):1493–9. Available from: http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.165.13.1493

44. Norman G. Research in clinical reasoning: past history and current trends. Med Educ [Internet]. 2005 Apr [cited 2018 Oct 11];39(4):418–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15

813765

45. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ [Internet]. 2005 Jan [cited 2018 Jun 11];39(1):98–106. Available from: http://doi.wiley.com/10.1111/j.1365-2929.2004.01972.x

46. Rimoldi HJA, Devane J, Haley J. Characterization of Processes. Educ Psychol Meas [Internet]. 1961 Jun 2 [cited 2018 Oct 11];21(2):383–92. Available from: http://journals.sagepub.com/doi/10.1177/001316446102100212

47. McGuire CH, Babbott D. SIMULATION TECHNIQUE IN THE MEASUREMENT OF PROBLEM-SOLVING SKILLS1. J Educ Meas [Internet]. 1967 Mar 1 [cited 2018 Oct 11];4(1):1–10. Available from: http://doi.wiley.com/10.1111/j.1745-3984.1967.tb00562.x

48. Elstein AS (Arthur S. Medical problem solving : an analysis of clinical reasoning

[Internet]. [cited 2018 Oct 11]. 350 p. Available from: http://www.hup.harvard.edu/catalog.php?isbn=9780674189089

49. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. Med Educ [Internet]. 1981 Sep [cited 2018 Oct 11];15(5):315–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6973686

50. Chase WG, Simon HA. Perception in chess. Cogn Psychol [Internet]. 1973 Jan [cited 2018 Oct 11];4(1):55–81. Available from: http://linkinghub.elsevier.com/retrieve/pii/0010028573900042

51. Groot AD de. Thought and choice in chess [Internet]. [cited 2018 Oct 11]. 480 p. Available from: https://books.google.co.uk/books?hl=en&lr=&id=dURbDAAAQBAJ&oi=fnd&pg=PR5&dq=De+groot++Perception+and+memory+

versus+thought:&ots=iZWWYm8eDH&sig=R4PvKkL0vcJlkTCtn0JYgDKluL0#v=onepage&q=De groot Perception and memory versus thought%3A&f=false

52. Patel VL, Groen GJ. Knowledge Based Solution Strategies in Medical Reasoning. Cogn Sci [Internet]. 1986 Jan [cited 2018 Oct 11];10(1):91–116. Available from: http://doi.wiley.com/10.1207/s15516709cog1001_4

53. Norman GR, Brooks LR, Allen SW. Recall by expert medical practitioners and novices as a record of processing attention. J Exp Psychol Learn Mem Cogn [Internet]. 1989 [cited 2018 Oct 11];15(6):1166–74. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.15.6.1166

54. Verkoeijen PPJL, Rikers RMJP, Schmidt HG, van de Wiel MWJ, Kooman JP. Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical

context. Med Educ [Internet]. 2004 Jun [cited 2018 Oct 12];38(6):617–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15189258

55. Schmidt HG, Boshuizen HP. On the origin of intermediate effects in clinical case recall. Mem Cognit [Internet]. 1993 May [cited 2018 Oct 11];21(3):338–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8316096

56. Burns BD. The Effects of Speed on Skilled Chess Performance. Psychol Sci [Internet]. 2004 Jul 6 [cited 2018 Oct 12];15(7):442–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15200627

57. Barrows HS, Feltovich PJ. The clinical reasoning process. Med Educ [Internet]. 1987 Mar [cited 2018 Oct 12];21(2):86–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35

74170

58. Hobus, P. P. M., Hofstra, M. L., Boshuizen, H. P. A., and Schmidt HG. Mental Representation of Prototypical Patients: Expert-Novice Differences. Paper presented at the First European Congress on Psychology, Amsterdam, 1989.

59. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med [Internet]. 1991 Sep [cited 2018 Oct 12];66(9 Suppl):S70-2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1930535

60. Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. Acad Med [Internet]. 1997 Oct [cited 2018 Oct 12];72(10 Suppl 1):S37-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9347733

61. Monteiro S, Norman G, Sherbino J. The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. J Eval Clin Pract [Internet]. 2018 Jun [cited 2018 Jun 8];24(3):666–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29532584

62. Evans JSBT, Stanovich KE. Dual-Process Theories of Higher Cognition. Perspect Psychol Sci [Internet]. 2013 May 7 [cited 2018 Jun 8];8(3):223–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26172965

63. Norman G. Teaching basic science to optimize transfer. Med Teach [Internet]. 2009 Sep [cited 2018 Jun 11];31(9):807–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19811185

64. Patel VL, Groen GJ, Norman GR. Effects of conventional and problem-based medical curricula on problem solving. Acad Med

[Internet]. 1991 Jul [cited 2018 Jun 11];66(7):380–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2059263

65. Quilici JL, Mayer RE. Role of examples in how students learn to categorize statistics word problems. J Educ Psychol [Internet]. 1996 [cited 2018 Jun 11];88(1):144–61. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.88.1.144

66. Catrambone R, Holyoak KJ. Overcoming contextual limitations on problem-solving transfer. J Exp Psychol Learn Mem Cogn [Internet]. 1989 [cited 2018 Jun 11];15(6):1147–56. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.15.6.1147

67. Loewenstein J, Thompson L, Gentner D. Analogical encoding facilitates knowledge transfer in negotiation. Psychonomic Bull Rev [Internet]. 1999 [cited 2018 Jun 11];6(4):586–97. Available from:

https://pdfs.semanticscholar.org/320a/97f429100effea969cbdf8ced51fdb112fbc.pdf

68. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ [Internet]. 2009 Apr [cited 2018 Oct 25];43(4):303–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19335571

69. McGee JB, Neill J, Goldman L, Casey E. Using multimedia virtual patients to enhance the clinical curriculum for medical students. Stud Health Technol Inform [Internet]. 1998 [cited 2019 Mar 26];52 Pt 2:732–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10384556

70. Schmidt HG, Ilvia Mamede & S. How to improve the teaching of clinical reasoning: a narrative review and a proposal medical education in review INTRODUCTION. John Wiley Sons Ltd Med Educ [Internet]. 2015 [cited 2018 Oct 22];49:961–73. Available

from: www.mededuc.comdiscuss.

71. Kenny P, Parsons TD, Gratch J, Leuski A, Rizzo AA. Virtual Patients for Clinical Therapist Skills Training. In: Intelligent Virtual Agents [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007 [cited 2018 Oct 26]. p. 197–210. Available from: http://link.springer.com/10.1007/978-3-540-74997-4_19

72. Norman G, Dore K, Krebs J, Neville AJ. The Power of the Plural: Effect of Conceptual Analogies on Successful Transfer. Acad Med [Internet]. 2007 Oct [cited 2018 Oct 25];82(Suppl):S16–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17895680

73. Huwendiek S, De leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach [Internet]. 2009 Aug [cited 2018 Oct 25];31(8):743–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19

811212

74. Bearman M, Cesnik B, Liddell M. Random comparison of “virtual patient” models in the context of teaching clinical communication skills. Med Educ [Internet]. 2001 Sep [cited 2018 Oct 26];35(9):824–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11555219

75. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients - what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ [Internet]. 2015 Dec 1 [cited 2018 Jun 11];15(1):11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25638167

76. Huang G, Reynolds R, Candler C. Virtual Patient Simulation at U.S. and Canadian Medical Schools. Acad Med [Internet]. 2007 May [cited 2019 Mar 26];82(5):446–51. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/17457063

77. Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL. Multi-institutional development and utilization of a computer-assisted learning program for the pediatrics clerkship: the CLIPP Project. Acad Med [Internet]. 2005 Sep [cited 2018 Jun 11];80(9):847–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16123465

78. Schifferdecker KE, Berman NB, Fall LH, Fischer MR. Adoption of computer-assisted learning in medical education: the educators' perspective. Med Educ [Internet]. 2012 Nov [cited 2018 Jun 11];46(11):1063–73. Available from: http://doi.wiley.com/10.1111/j.1365-2923.2012.04350.x

79. Virtual Reality Market Share & Trends Report, 2021-2028 [Internet]. [cited 2021 Nov 30]. Available from:

https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market

80. Andrew Ward. Cutting edge virtual reality streams live operation to trainees | Financial Times [Internet]. Financial Times. 2016 [cited 2018 Oct 26]. Available from: https://www.ft.com/content/a375879c-0215-11e6-99cb-83242733f755

81. Facebook gives a glimpse of metaverse, its planned virtual reality world – video | Technology | The Guardian [Internet]. [cited 2021 Nov 30]. Available from: https://www.theguardian.com/technology/video/2021/oct/29/facebook-gives-a-glimpse-of-metaverse-its-planned-virtual-reality-world-video

82. Sherman WR, Craig AB. Understanding Virtual Reality: Interface, Application, and Design [Internet]. [cited 2018 Oct 27]. Available from: https://profs.info.uaic.ro/~avitcu/FII 2015-2016/Animatie 3D_Documentatie/VR.pdf

83. Zhang Y, Song H, Liu X, Tang D, Chen YE,

Zhang X. Language learning enhanced by massive multiple online Role-Playing games (MMORPGs) and the underlying behavioral and neural mechanisms. Vol. 11, Frontiers in Human Neuroscience. 2017.

84. Melús-Palazón E, Bartolomé-Moreno C, Palacín-Arbués JC, Lafuente-Lafuente A, García IG, Guillen S, et al. Experience with using second life for medical education in a family and community medicine education unit. BMC Med Educ. 2012;12(1):1–9.

85. Rupp MA, Kozachuk J, Michaelis JR, Odette KL, Smither JA, McConnell DS. The effects of immersiveness and future VR expectations on subjec-tive-experiences during an educational 360° video. Proc Hum Factors Ergon Soc Annu Meet [Internet]. 2016 Sep 8 [cited 2019 Feb 6];60(1):2108–12. Available from: http://journals.sagepub.com/doi/10.1177/1541931213601477

86. Sheridan TB. Musings on Telepresence and Virtual Presence. Presence Teleoperators Virtual Environ. 1992 Jan;1(1):120–6.

87. Witmer BG, Singer MJ. Measuring presence in virtual environments: A presence questionnaire. Presence Teleoperators Virtual Environ [Internet]. 1998 [cited 2021 Mar 24];7(3):225–40. Available from: /record/2016-00465-001

88. Sas C, O'Hare GMP. Presence Equation: An Investigation into Cognitive Factors Underlying Presence. Vol. 12, Presence: Teleoperators and Virtual Environments. 2003. p. 523–37.

89. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence Teleoperators Virtual Environ [Internet]. 1997 Dec 1 [cited 2021 Mar 24];6(6):603–16. Available from: https://direct.mit.edu/pvar/article/6/6/60

3-616/18157

90. Bailey JH, Witmer BG. Learning and Transfer of Spatial Knowledge in a Virtual Environment. Proc Hum Factors Ergon Soc Annu Meet [Internet]. 1994 Oct 5 [cited 2021 Mar 24];38(18):1158–62. Available from: http://journals.sagepub.com/doi/10.1177/154193129403801803

91. Papadakis G, Mania K, Coxon M, Koutroulis E. The effect of tracking delay on awareness states in immersive Virtual Environments: An initial exploration. In: Proceedings of VRCAI 2011: ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications to Industry [Internet]. New York, New York, USA: ACM Press; 2011 [cited 2021 Mar 24]. p. 475–82. Available from: http://dl.acm.org/citation.cfm?doid=2087756.2087848

92. Slater M, Linakis V, Usoh M, Kooper R. Immersion, presence and performance in

virtual environments. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology - VRST ’96 [Internet]. New York, New York, USA: Association for Computing Machinery (ACM); 1996 [cited 2021 Mar 24]. p. 163–72. Available from: http://dl.acm.org/citation.cfm?doid=3304181.3304216

93. Dede C. Immersive interfaces for engagement and learning [Internet]. Vol. 323, Science. American Association for the Advancement of Science; 2009 [cited 2021 Mar 24]. p. 66–9. Available from: https://science.sciencemag.org/content/323/5910/66

94. O_brien HL, Toms EG. What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology 1. 2008 [cited 2021 Mar 24]; Available from: http://www.asis.org/jasist.html.

95. Tang KS, Cheng DL, Mi E, Greenberg PB.

Augmented reality in medical education: a systematic review. Can Med Educ J [Internet]. 2020 Dec 6 [cited 2022 Jun 10];11(1):e81. Available from: /pmc/articles/PMC7082471/

96. Tursø-Finnich T, Jensen RO, Jensen LX, Konge L, Thinggaard E. Virtual Reality Head-Mounted Displays in Medical Education—A Systematic Review. Simul Healthc J Soc Simul Healthc. 2022;Publish Ah(00):1–9.

97. Harrington CM, Kavanagh DO, Quinlan JF, Ryan D, Dicker P, O'Keeffe D, et al. Development and evaluation of a trauma decision-making simulator in Oculus virtual reality. Am J Surg [Internet]. 2018 Jan 1 [cited 2018 Oct 26];215(1):42–7. Available from: https://www.sciencedirect.com/science/article/pii/S0002961017300405#bib23

98. Lowe J, Peng C, Winstead-Derlega C, Curtis H. 360 virtual reality pediatric mass casualty incident: A cross sectional

observational study of triage and out-of-hospital intervention accuracy at a national conference. J Am Coll Emerg Physicians Open [Internet]. 2020 Oct [cited 2022 Jun 9];1(5):974. Available from: /pmc/articles/PMC7593497/

99. Aussedat C, Robier M, Aoustin JM, Parietti-Winkler C, Lescanne E, Bonnard D, et al. Using virtual reality in audiological training: Our experience in 22 otolaryngology residents. Clin Otolaryngol [Internet]. 2020 Jul 1 [cited 2022 Jun 9];45(4):643–8. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/coa.13554

100. Zackoff MW, Real FJ, Sahay RD, Fei L, Guiot A, Lehmann C, et al. Impact of an Immersive Virtual Reality Curriculum on Medical Students' Clinical Assessment of Infants with Respiratory Distress. Pediatr Crit Care Med. 2020;477–85.

101. Rheingold H. Virtual reality [Internet]. Summit Books; 1991 [cited 2018 Oct 27].

415 p. Available from: http://www.simonandschuster.com/books/Virtual-Reality/Howard-Rheingold/9780671778972

102. Rosson L. The Virtual Interface Environment Workstation (VIEW), 1990. 2014 [cited 2018 Oct 28]; Available from: https://www.nasa.gov/ames/spinoff/new_continent_of_ideas/

103. Paul Sorene. Jaron Lanier's EyePhone: Head And Glove Virtual Reality In The 1980s - Flashbak [Internet]. [cited 2018 Oct 28]. Available from: https://flashbak.com/jaron-laniers-eyephone-head-and-glove-virtual-reality-in-the-1980s-26180/

104. Elyse Betters. Virtual reality: Lessons from the past for Oculus Rift - BBC News [Internet]. [cited 2018 Oct 28]. Available from: https://www.bbc.co.uk/news/technology-23877695

105. Luke Dormehl. 8 Major Milestones in

the Brief History of Virtual Reality | Digital Trends [Internet]. [cited 2018 Oct 28]. Available from: https://www.digitaltrends.com/cool-tech/history-of-virtual-reality/

106. Cornwell BR, Johnson L, Berardi L, Grillon C. Anticipation of Public Speaking in Virtual Reality Reveals a Relationship Between Trait Social Anxiety and Startle Reactivity. Biol Psychiatry [Internet]. 2006 Apr 1 [cited 2018 Oct 26];59(7):664–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16325155

107. Krijn M, Emmelkamp PM., Biemond R, de Wilde de Ligny C, Schuemie MJ, van der Mast CAP. Treatment of acrophobia in virtual reality: The role of immersion and presence. Behav Res Ther [Internet]. 2004 Feb [cited 2018 Oct 26];42(2):229–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14975783

108. Botella C, Serrano B, baños R, García-Palacios A. Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: a review of its efficacy, the adequacy of the treatment protocol, and its acceptability. Neuropsychiatr Dis Treat [Internet]. 2015 Oct [cited 2018 Oct 26];11:2533. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26491332

109. Hoffman HG, Patterson DR, Seibel E, Soltani M, Jewett-Leahy L, Sharar SR. Virtual Reality Pain Control During Burn Wound Debridement in the Hydrotank. Clin J Pain [Internet]. 2008 May [cited 2018 Oct 26];24(4):299–304. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18427228

110. Chirico A, Lucidi F, De Laurentiis M, Milanese C, Napoli A, Giordano A. Virtual Reality in Health System: Beyond

Entertainment. A Mini-Review on the Efficacy of VR During Cancer Treatment. J Cell Physiol [Internet]. 2016 Feb [cited 2018 Oct 26];231(2):275–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26238976

111. Bekelis K, Calnan D, Simmons N, MacKenzie TA, Kakoulides G. Effect of an Immersive Preoperative Virtual Reality Experience on Patient Reported Outcomes. Ann Surg [Internet]. 2017 Jun [cited 2018 Oct 26];265(6):1068–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27906757

112. Boyd LE, Day K, Stewart N, Abdo K, Lamkin K, Linstead E. Leveling the Playing Field: Supporting Neurodiversity Via Virtual Realities. Technol Innov. 2019;20(1):105–16.

113. Maskati E, Alkeraiem F, Khalil N, Baik R, Aljuhani R, Alsobhi A. Using Virtual

Reality (VR) in Teaching Students With Dyslexia. Int J Emerg Technol Learn. 2021;16(9):291–305.

114. Rizzo A “Skip”, Kim GJ. A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. Presence Teleoperators Virtual Environ [Internet]. 2005 Apr 13 [cited 2018 Oct 26];14(2):119–46. Available from: http://www.mitpressjournals.org/doi/10.1162/1054746053967094

115. Elyse Wanshel. Virtual Reality Lets Med Students Experience What It’s Like To Be 74 | HuffPost UK [Internet]. [cited 2018 Oct 26]. Available from: https://www.huffingtonpost.co.uk/entry/we-are-alfred-embodied-labs-carrie-shaw-virtual-reality-medical-students-elderly-geriatric-care_us_57505bbce4b0c3752dccbeaa?guccounter=1&guce_referrer_us=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvLnVrLw&guce_referrer_cs=vzXAzOkZUH

116. Huang G, Reynolds R, Candler C. Virtual Patient Simulation at U.S. and Canadian Medical Schools. Acad Med [Internet]. 2007 May [cited 2020 Apr 6];82(5):446–51. Available from: http://journals.lww.com/00001888-200705000-00004

117. Posel N, Fleiszer D, Shore BM. 12 Tips: Guidelines for authoring virtual patient cases. Med Teach [Internet]. 2009 Jan 9 [cited 2019 Feb 13];31(8):701–8. Available from: http://www.tandfonline.com/doi/full/10.1080/01421590902793867

118. Scaramuzza D. Omnidirectional Camera [Internet]. [cited 2019 Jan 3]. Available from: http://rpg.ifi.uzh.ch/docs/omnidirectional_camera.pdf

119. Ho T, Budagavi M. DUAL-FISHEYE LENS STITCHING FOR 360-DEGREE IMAGING [Internet]. 2017 [cited 2019 Jan 3]. Available from:

https://arxiv.org/pdf/1708.08988.pdf

120. Huang J, Chen Z, Ceylan D, Jin H. 6-DOF VR videos with a single 360-camera. In: 2017 IEEE Virtual Reality (VR) [Internet]. IEEE; 2017 [cited 2019 Jan 2]. p. 37–44. Available from: http://ieeexplore.ieee.org/document/7892229/

121. Google Cardboard – Google VR [Internet]. [cited 2021 Mar 24]. Available from: https://arvr.google.com/cardboard/

122. (No Title) [Internet]. [cited 2021 Mar 24]. Available from: https://blog.kickpush.co/beyond-reality-first-steps-into-the-unknown-cbb19f039e51?gi=7a2e90cd3891

123. Ou L, Chen J, Hillman K, Flabouris A, Parr M, Assareh H, et al. The impact of post-operative sepsis on mortality after hospital discharge among elective surgical patients: a population-based cohort study. Crit Care [Internet]. 2017 Dec 20 [cited 2020 Feb 21];21(1):34. Available from:

http://ccforum.biomedcentral.com/articles/10.1186/s13054-016-1596-7

124. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med [Internet]. 2009 Oct 1 [cited 2020 Feb 21];361(14):1368–75. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMsa0903048

125. Round J, Conradi E, Poulton T. Improving assessment with virtual patients. Med Teach [Internet]. 2009 Aug [cited 2019 Feb 4];31(8):759–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19811215

126. CLAUSER BE, MARGOLIS MJ, SWANSON DB. An Examination of the Contribution of Computer-based Case Simulations to the USMLE Step 3 Examination. Acad Med [Internet]. 2002 Oct [cited 2019 Feb 4];77(Supplement):S80–2. Available from:

https://insights.ovid.com/crossref?an=00001888-200210001-00026

127. Fors UGH, Gunning WT. The Impact of Different Scoring Rubrics for Grading Virtual Patient-Based Exams. J Educ Comput Res [Internet]. 2014 Jan 22 [cited 2019 Feb 4];50(1):97–118. Available from: http://journals.sagepub.com/doi/10.2190/EC.50.1.e

128. Brooke J. SUS-A quick and dirty usability scale [Internet]. [cited 2019 Jan 21]. Available from: https://hell.meiert.org/core/pdf/sus.pdf

129. LaViola JJ. A discussion of cybersickness in virtual environments. ACM SIGCHI Bull. 2000 Jan;32(1):47–56.

130. Schreuder HWR, Van Dongen KW, Roeleveld SJ, Schijven MP, Broeders IAMJ. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. 2009 [cited 2019 Jan 21]; Available from: www.AJOG.org

131. Sharples S, Cobb S, Moody A, Wilson JR. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. Displays [Internet]. 2008 Mar [cited 2021 Mar 24];29(2):58–69. Available from: /record/2008-00055-001

132. Slater M. How colorful was your day? Why questionnaires cannot assess presence in virtual environments [Internet]. Vol. 13, Presence: Teleoperators and Virtual Environments. MIT Press; 2004 [cited 2021 Mar 24]. p. 484–93. Available from: http://direct.mit.edu/pvar/article-pdf/13/4/484/1624147/1054746041944849.pdf

133. Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. Presence Teleoperators Virtual Environ. 2000;9(5):497–503.

134. Slater M, Sadagic A, Usoh M,

Schroeder R. Small-group behavior in a virtual and real environment: A comparative study. Presence Teleoperators Virtual Environ [Internet]. 2000 Feb 1 [cited 2021 Mar 24];9(1):37–51. Available from: https://direct.mit.edu/pvar/article/9/1/37-51/18271

135. Kim H, Kim DJ, Chung WH, Park KA, Kim JDK, Kim D, et al. Clinical predictors of cybersickness in virtual reality (VR) among highly stressed people. Sci Reports 2021 111 [Internet]. 2021 Jun 9 [cited 2021 Nov 30];11(1):1–11. Available from: https://www.nature.com/articles/s41598-021-91573-w

136. Cobb SVG, Nichols S, Ramsey A, Wilson JR. Virtual reality-induced symptoms and effects (VRISE). Presence Teleoperators Virtual Environ. 1999;8(2):169–86.

137. Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in

recognizing one’s own incompetence lead to inflated self-assessments. J Pers Soc Psychol [Internet]. 1999 Dec [cited 2021 Nov 30];77(6):1121–34. Available from: /record/1999-15054-002

138. Newstrom JW, Rubenfeld SA. The Johari Window: A Reconceptualization. Dev Bus Simul Exp Learn Proc Annu ABSEL Conf [Internet]. 1983 Mar 13 [cited 2021 Nov 30];10. Available from: https://absel-ojs-ttu.tdl.org/absel/index.php/absel/article/view/2298

139. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: From theory to practice: AMEE Guide No. 75. Med Teach [Internet]. 2013 Mar 29 [cited 2019 Feb 21];35(3):184–93. Available from: http://www.tandfonline.com/doi/full/10.3109/0142159X.2013.760036

140. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for

construction. BMC Med Inform Decis Mak [Internet]. 2008 May 6 [cited 2019 Feb 21];8:18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18460199

141. Charlin B, van der Vleuten C. Standardized Assessment of Reasoning in Contexts of Uncertainty. Eval Health Prof [Internet]. 2004 Sep 24 [cited 2019 Mar 21];27(3):304–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15312287

142. Gaba DM. Improving anesthesiologists' performance by simulating reality. Anesthesiology [Internet]. 1992 Apr 1 [cited 2021 Nov 30];76(4):491–4. Available from: https://europepmc.org/article/med/1550272

143. Flanagan B, Nestel D, Joseph M. Making patient safety the focus: Crisis Resource Management in the undergraduate curriculum. Med Educ

[Internet]. 2004 Jan 1 [cited 2021 Nov 30];38(1):56–66. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2923.2004.01701.x

144. Krishnan D, Keloth AV, Ubedulla S. Pros and cons of simulation in medical education: A review. undefined. 2017;

145. McFetrich J. A structured literature review on the use of high fidelity patient simulators for teaching in emergency medicine. Emerg Med J [Internet]. 2006 Jul [cited 2021 Nov 30];23(7):509. Available from: /pmc/articles/PMC2579539/

146. Kalyuga S, Ayres P, Chandler P, Sweller J. The Expertise Reversal Effect. https://doi.org/101207/S15326985EP3801_4 [Internet]. 2010 [cited 2021 Nov 30];38(1):23–31. Available from: https://www.tandfonline.com/doi/abs/10.1207/S15326985EP3801_4

147. Vroom VH (Victor H. Work and motivation. New York: Wiley; 1964. 331 p.

148. Raemer D, Anderson M, Cheng A, Fanning R, Nadkarni V, Savoldelli G. Research Regarding Debriefing as Part of the Learning Process. Simul Healthc J Soc Simul Healthc [Internet]. 2011 Aug [cited 2018 Jun 8];6:S52–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21817862

149. Prince KJAH, Boshuizen HPA, van der Vleuten CPM, Scherpbier AJJA. Students' opinions about their preparation for clinical practice. Med Educ [Internet]. 2005 Jul [cited 2018 Jun 19];39(7):704–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15960791

150. Bellini LM, Baime M, Shea JA. Variation of mood and empathy during internship. JAMA [Internet]. 2002 Jun 19 [cited 2018 Jun 19];287(23):3143–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12069680

151. Brennan N, Corrigan O, Allard J, Archer J, Barnes R, Bleakley A, et al. The transition from medical student to junior doctor: today's experiences of Tomorrow's Doctors. Med Educ [Internet]. 2010 May [cited 2018 Jun 19];44(5):449–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20518984

152. Brown JM, Ryland I, Shaw NJ, Graham DR. Working as a newly appointed consultant: a study into the transition from specialist registrar. Br J Hosp Med (Lond) [Internet]. 2009 Jul [cited 2018 Jun 19];70(7):410–4. Available from: http://www.magonlinelibrary.com/doi/10.12968/hmed.2009.70.7.43126

153. Brown J, Chapman T, Graham D. Becoming a new doctor: a learning or survival exercise? Med Educ [Internet]. 2007 Jul [cited 2018 Jun 19];41(7):653–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17

614885

154. Hurst C, Kahan D, Ruetalo M, Edwards S. A year in transition: a qualitative study examining the trajectory of first year residents' well-being. BMC Med Educ [Internet]. 2013 Jul 10 [cited 2018 Jun 19];13(1):96. Available from: http://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-13-96

155. Liu C-H, Tang W-R, Weng W-H, Lin Y-H, Chen C-Y. The process of coping with stress by Taiwanese medical interns: a qualitative study. BMC Med Educ [Internet]. 2016 Jan 12 [cited 2018 Jun 19];16(1):10. Available from: http://www.biomedcentral.com/1472-6920/16/10

156. Morrow G, Burford B, Redfern N, Briel R, Illing J. Does specialty training prepare doctors for senior roles? A questionnaire study of new UK consultants. Postgrad Med J [Internet]. 2012 Oct [cited 2018 Jun 19];88(1044):558–65. Available from:

http://pmj.bmj.com/lookup/doi/10.1136/postgradmedj-2011-130460

157. Tallentire VR, Smith SE, Skinner J, Cameron HS. Understanding the behaviour of newly qualified doctors in acute care contexts. Med Educ [Internet]. 2011 Oct [cited 2018 Jun 19];45(10):995–1005. Available from: http://doi.wiley.com/10.1111/j.1365-2923.2011.04024.x

158. Ker J, Hesketh A, Anderson F, Johnston D. PRHO views on the usefulness of a pilot ward simulation exercise. Hosp Med [Internet]. 2005 Mar [cited 2018 Jun 20];66(3):168–70. Available from: http://www.magonlinelibrary.com/doi/10.12968/hmed.2005.66.3.17687

159. Ker JS, Hesketh EA, Anderson F, Johnston DA. Can a ward simulation exercise achieve the realism that reflects the complexity of everyday practice junior doctors encounter? Med Teach [Internet]. 2006 Jun 3 [cited 2018 Jun 20];28(4):330–

4. Available from: http://www.tandfonline.com/doi/full/10.1080/01421590600627623

160. Antonoff MB, Swanson JA, Green CA, Mann BD, Maddaus MA, D'Cunha J. The Significant Impact of a Competency-Based Preparatory Course for Senior Medical Students Entering Surgical Residency. Acad Med [Internet]. 2012 Mar [cited 2018 Jun 20];87(3):308–19. Available from: https://insights.ovid.com/crossref?an=00001888-201203010-00017

Appendices

Appendix 1 - Case 1 (Mr. Prasad)

1/4

Admission Clerking

PC: Abdominal pain

HPC: 65 year old male 48-hour colicky lower abdominal pain, abdominal distension with vomiting, direct questioning - anorexic for 1/52 and constipated for 1/52. Lethargic for 3/12.

DD - ?bowel obstruction – likely malignancy based on age / anorexia / lethargy

2/4

PMH: HTN, hypercholesterolaemia, unstable angina - due for elective coronary angiogram, one year post-TKR for OA.

Drugs: Ramipril 5 mg, bendroflumethiazide 2.5 mg, atorvastatin 10 mg nocte, ferrous sulphate 200 mg TDS, GTN PRN, 75 mg aspirin. NKDA.

Medical hx may necessitate going to HDU,

Due to unstable angina and plan for angio – significantly increases risk - ?need for echo / cardiology review, pre-op optimisation

3/4

FH: COPD in Father, died aged 86 years.

SH: Retired - former financial adviser. Current smoker - 2 cigarettes per day, pack-year: 5. Teetotaller, no substance abuse. Lives in 2-storey flat with wife. Independent in ADLs.

4/4

O/E –
Resp - Chest clear,
CVS - I + II + 0,
Abdo: distended abdomen, generalised tenderness, no groin masses

Plan:

1. Bloods for FBC, U&Es, CRP, VBG, Coag, G&S,
2. Fluids: Nil by mouth - discuss CT with consultant

1/2
Daily Notes:

D1 – well – obs stable – on 2L O2 n.c. – in pain overnight – given analgesia as required – slept – high NG output – 600 mls overnight.

D2 – well – obs stable – on 2L O2 n.c. - remains taccy – sleeping – pain team have started a PCA – remain NBM – no bowel sounds – NG output – 200mls / 4 hrly – 1200ml/d. NG tube pulled out by patient

2/2

D3 – obs stable – on 2L O2 n.c. - remains taccy – reporting pain worsening – started drinking 30 ml/hr

1/2
Operation Notes

Urinary catheter inserted pre-op, WHO, abx, VTE prophylaxis (subcut heparin / TEDS)

Lower midline incision, mass in caecum, dilated loops of small bowel, caecum mobilised, right ureter seen and preserved, standard right hemicolectomy, side to side stapled anastomosis, mesenteric defect closed.

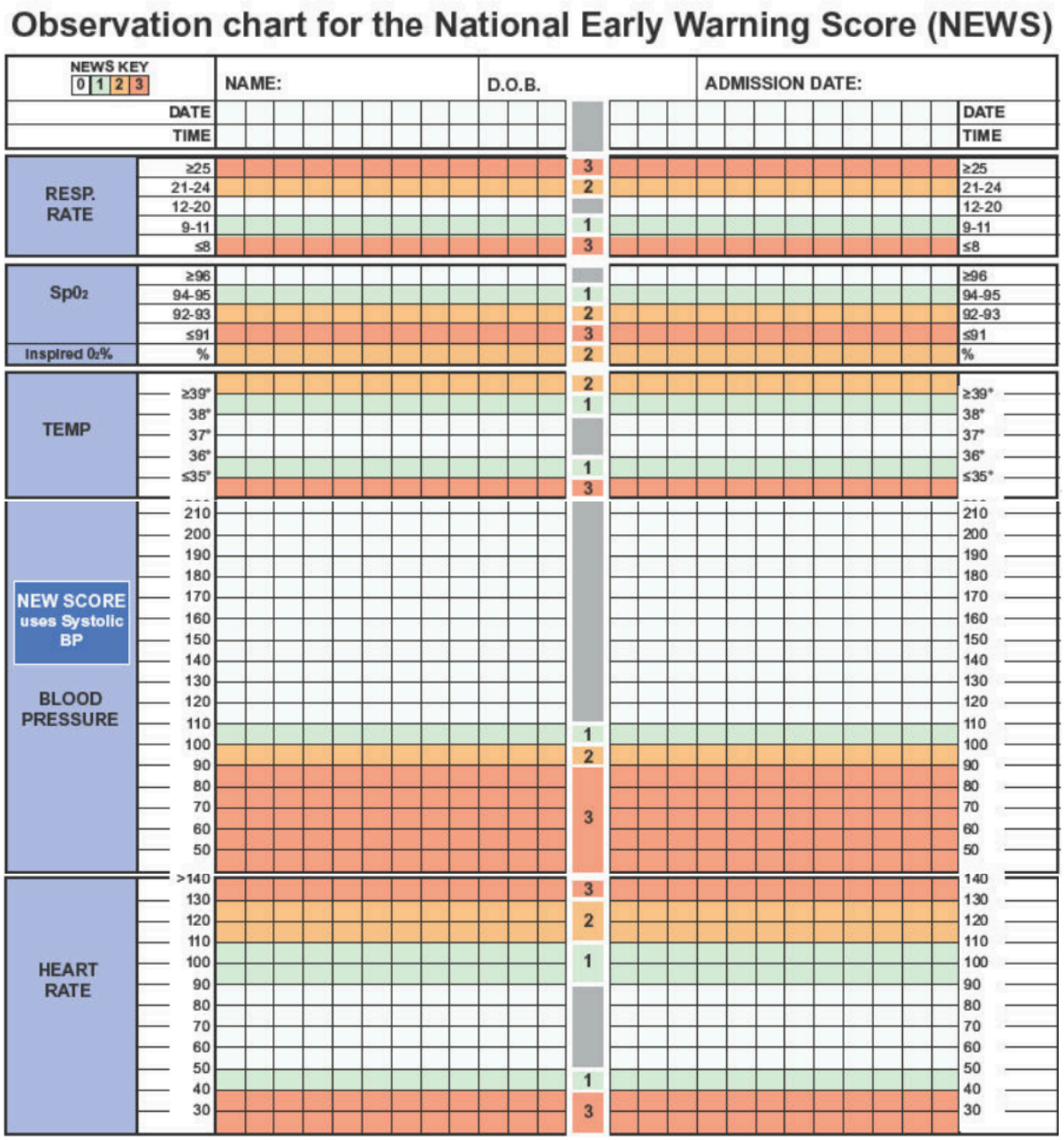
2/2

Closure – continuous loop nylon, staples to skin, op-site dressing.

Post-op instructions:

- 1. Chase histology
- 2. Routine obs
- 3. HDU bed

Obs 1:

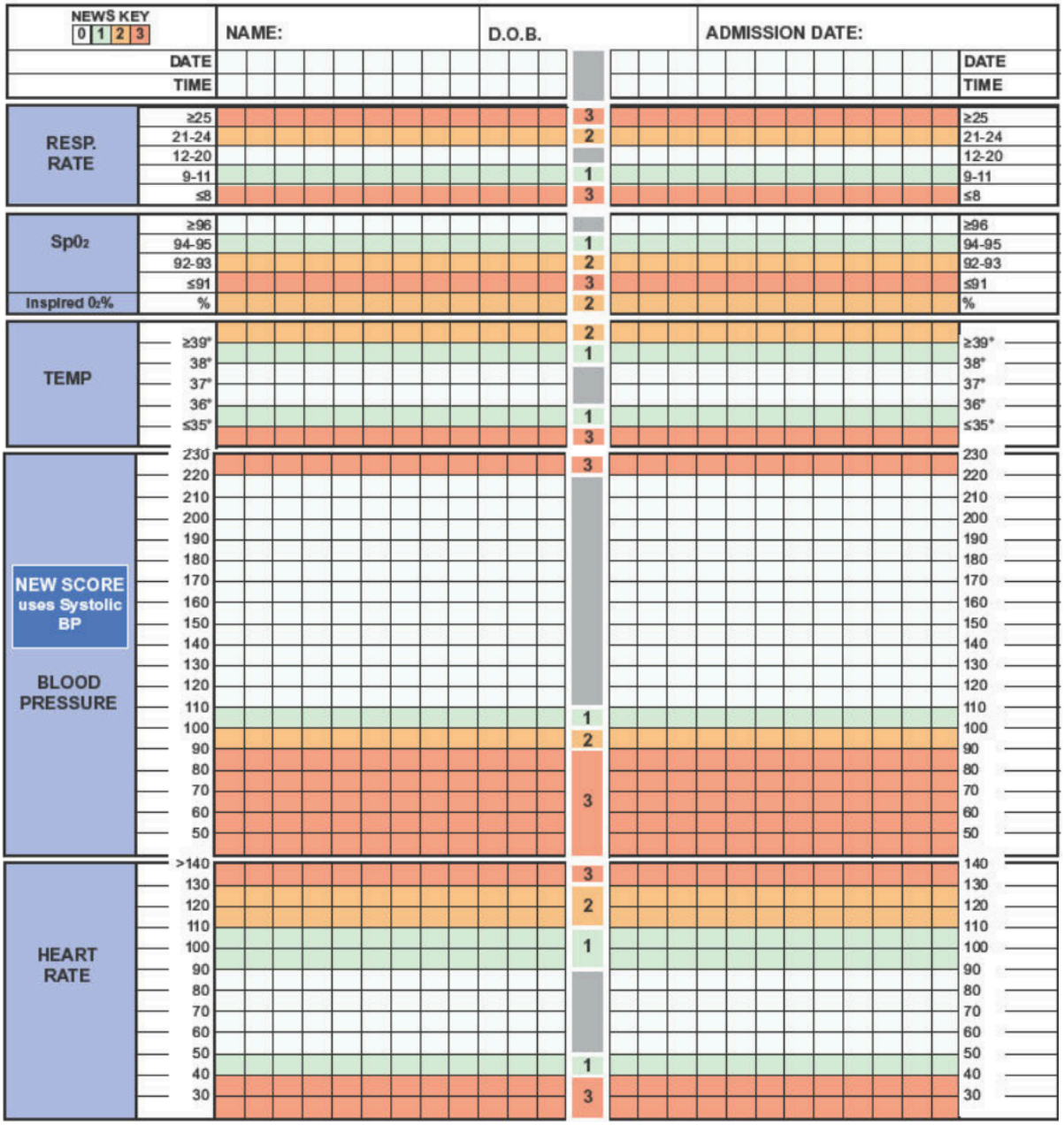


Please note, the obs chart is split into three parts so that high resolution screen shots can be taken when zoomed in. Please add copies of this marker ● over image or list each trend.

Info given:

- SIRS response D0 / D1 – 37.5 (below)
- Persistent sinus taccy – 105 bpm (100-110)
- RR – 12 constant
- BP – 110/70 immediately post-op
- 160/100mmHg now
- Obs 2: D2 – T 37.3, taccy 108, RR – 16, BP 160/100mmHg, Sats 99 on 2L
- D3 – T – 37.9, taccy 112, RR – 16, BP 145/90mmHg, Sat 99 on 2L

Observation chart for the National Early Warning Score (NEWS)



Fluid Prescription Chart

Date	Fluid	Vol. (mL)	Additive name	Additive dose	Time (hours)	Route	Sig.
Pre-op	Plasmolyte	1L		8hrs			
D0	Plasmolyte	1L		8 Hrs			
D1	Plasmolyte	1L		8 Hrs			
D2	NaCL	1L	40mmol K	8 Hrs			
D3	Plasmolyte	1L		8 Hrs			

^information given: '8-hourly bags'

Admission VBG:

Parameter	Result	Units	Reference Interval
pH	7.35	-	7.35-7.45
pCO2	6.2	kPa	4.7-6.0
pO2	10.5	kPa	10.5-14.0

HCO3-	20	mmol/L	22-26
Lactate	2.2	mmol/L	0.5-2.2
Potassium	3.5	mmol/L	3.3-5.3
Base excess	-2.5	mmol/L	-2 to +2 mEq/L

Urine:

Parameter	Result
Glucose	-ve
Leukocytes	++
Erythrocytes	-ve
Nitrites	-ve
Ketones	++

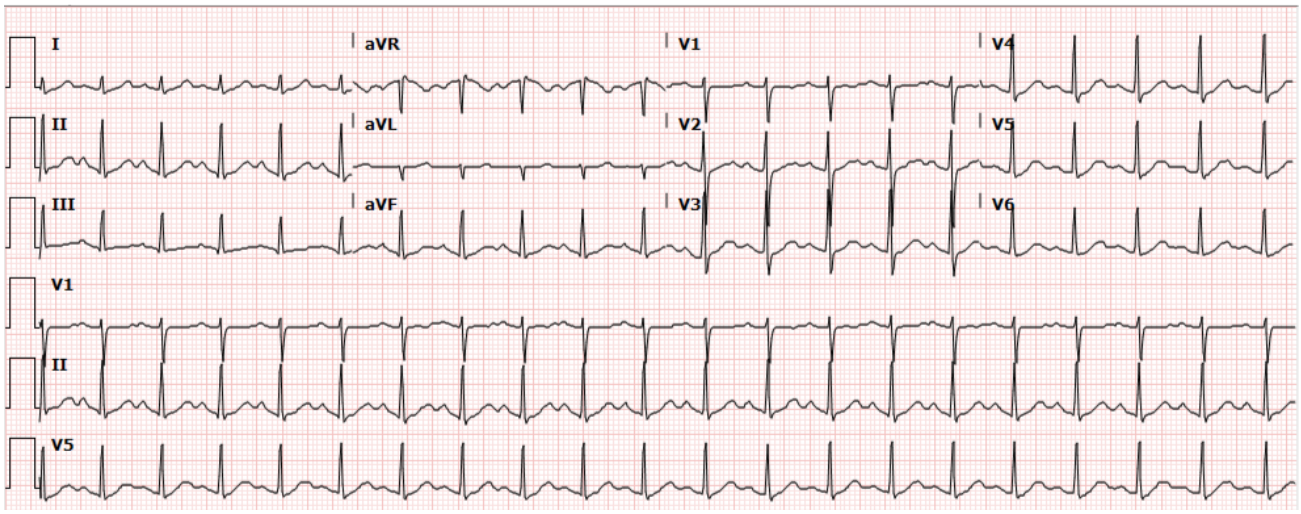
D3 ABG on 2L

Parameter	Result	Units	Reference Interval
pH	7.30	-	7.35-7.45

pCO2	6.2	kPa	4.7-6.0
pO2	9.6	kPa	10.5-14.0
HCO3-	22	mmol/L	22-26
Lactate	1.4	mmol/L	0.5-2.2
Potassium	3.0	mmol/L	3.3-5.3
Base excess	-1.9	mmol/L	-2 to +2 mEq/L

Say – needs more K in fluid

ECG: slower ECG with



Parameter	Result	Units	Reference Interval
WBC count	12	$X10^9/L$	4.5-11.0
Platelets	360	$x10^9/L$	140-400

Hb	89	g/L	115-165
Hematocrit	0.40		
MCV	76	fL	80-100
MCH	30	pg	27-32
MCHC	34.0	g/dL	32.0-36.0

D1 FBC:

Parameter	Result	Units	Reference Interval
WBC count	14	$X10^9/L$	4.5-11.0
Platelets	380	$x10^9/L$	140-400
Hb	96	g/L	115-165
Hematocrit	0.3	L/L	3.8-5.8
MCV	76	fL	80-100
MCH	30	pg	27-32
MCHC	34.0	g/dL	32.0-36.0
RDW	13.4	%	11.7-15.0

D3 FBC:

Parameter	Result	Units	Reference Interval
WBC count	17	$X10^9/L$	4.5-11.0
Platelets	380	$x10^9/L$	140-400
Hb	96	g/L	115-165
Hematocrit	0.4	L/L	3.8-5.8
MCV	76	fL	80-100
MCH	30	pg	27-32
MCHC	34.0	g/dL	32.0-36.0
RDW	13.4	%	11.7-15.0

On admission

LFTs:

Parameter	Result	Units	Reference Interval
Total bilirubin	15	µmol/L	<20
ALT	20	IU/L	<33
AST	24	IU/L	<45
ALP	86	IU/L	30-130
Gamma GT	31	IU/L	<40
Albumin	30	g/L	35-50

U&Es O/A

Parameter	Result	Units	Reference Interval
Creatinine	84	µmol/L	45-84
Na	135	mmol/L	133-146
K	3.5	mmol/L	3.5-5.3
Glucose	5.2	mmol/L	3.9-5.5

U&Es D1

Parameter	Result	Units	Reference Interval
Creatinine	90	µmol/L	45-84
Na	137	mmol/L	133-146
K	3.4	mmol/L	3.5-5.3
Glucose	5.2	mmol/L	3.9-5.5

U&Es D3

Parameter	Result	Units	Reference Interval
Creatinine	120	µmol/L	45-84
Na	135	mmol/L	133-146
K	3.0	mmol/L	3.5-5.3
Glucose	5.2	mmol/L	3.9-5.5

On admission

Parameter	Result	Units	Reference Interval
CRP	80	mg/L	<10

D3

Parameter	Result	Units	Reference Interval
CRP	145	mg/L	<10

Coags:

Parameter	Result	Units	Reference Interval
PT	12.0	seconds	11-14
INR	1.1	-	
APTT	27.0	seconds	24-37
TT	15.5	seconds	12.5-17.0

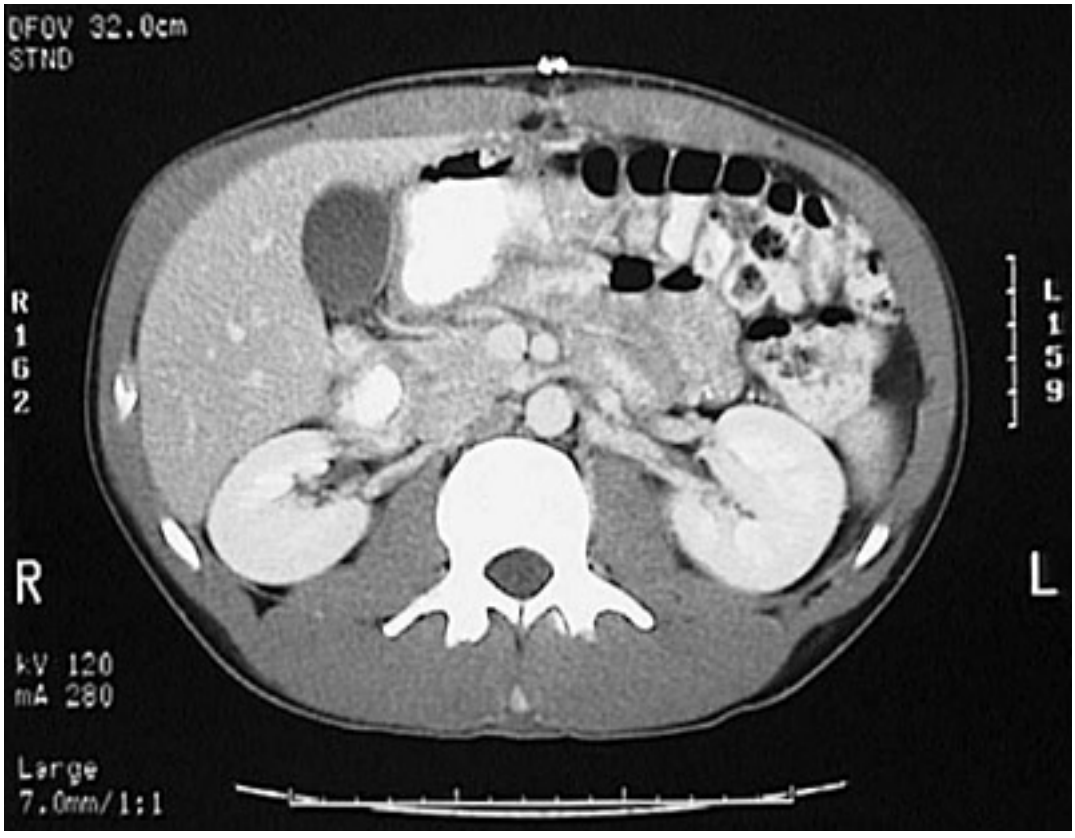
Fibrinogen	2.0	seconds	1.5-2.5

Please give values for:

CBG -

Cardiac biomarkers -

CT:



Misc info given:

- Stressors:
- Plan:
- Abx – HAP – protocol
- Fluids – increase iv fluids
- Ordering CXR
- ECG – sinus taccy

Appendix 2 Case 2 (Ms. Smith)

Handover from nursing staff

Handover from clinical staff

Hello doctor, thank you for coming down. This is Ms. Smith, she is 30 years old and a few days ago she underwent a bowel operation. She has been spiking temperatures, for which I have given her paracetamol, but they seem to persist. I need to go and see another patient and will be with you when finished.

History

Pain history – this pain started in the morning in my stomach. I can't keep anything down and I'm not feeling very well. What's wrong with me? (after time delay – why are you just standing there, don't you want to ask me any more questions?)

Drug history - I was on lots of medications beforehand, I think they have changed them. Can't you just look at the drug chart?

After time delay – are you actually going to do something or just stand there?

Ask about wound – the nurses changed the dressing yesterday and I did notice anything abnormal, but I'm sure they have written something in my notes.

After time delay – is there anything else you would like to ask me?

Family history – I can't think of anything. My grandma had a heart attack when she was 70, but is this question even relevant?

After time delay – do you want to ask me anything else?

Recent travel history – Are you actually serious? Is this really relevant to my current circumstance, I am in a huge amount of pain, can you please help me?

Bowel movements – I was passing flatus since normally since yesterday, today I haven't. I have not opened my bowels since the operation.

Anything else you want to ask me that's actually relevant?

Social history – I am engineer and I need to get back to my job soon. Are you actually going to do something or will you continue to ask irrelevant questions?

Operation note

Ms. Adrienne Smith

1/6/1987

MRN 6340897

Operating surgeon: Mr. B Colon

Assisting Surgeon: Mr. M. Barrett

Anaesthetist: Dr. K. Gas

Catheter inserted on table

Ileo-caecal resection – laparoscopic – converted to open due to difficulties with anastomosis.

GA, standard laparoscopic portals

Ileum mobilised without difficulty

Hepatic flexure mobilised – some bleeding encountered – controlled with haemostat gel

Splenic flexure mobilised without difficulty

Resection from distal ileum to descending colon

Staple gun used – end prepared – failure to deploy necessitating conversion to open procedure. Prepared manually – another staple gun found – finally completed with gun.

Midline wound closed in layers – 1 Vicryl / 2-0 Vicryl / 3-0 Monocryl

Opsite on all wounds

Post-op plan:

1. PCA
2. Sit up this evening
3. Enhance recovery protocol

Mr. M. Barrett

Admission clerking

PC:

Elective admission for surgical mx of Crohn's disease

HPC:

Five year history of Crohn's disease managed with Methotrexate. She has recently had a worsening of symptoms – unresponsive to steroid treatment. Decision made to list for ileo-caecal resection based on severity of symptoms and failure of medical treatment.

Sx's have been stable for the last few days – BO around x2/d, no blood, painfree currently

PMH:

Nil else

DH:

Penicillin allergy

Azathioprine – 100mg PO OD

Prednisolone – 20 mg PO OD

MTX – 25 mg PO weekly

SH:

Aerospace engineer

Lives with partner in a ground floor apartment

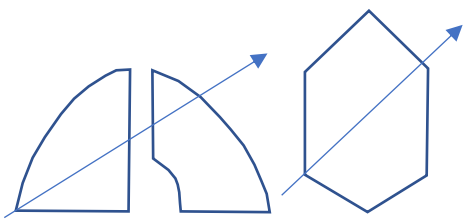
Non-smoker

Etoh – 8 units/week

FH:

Nil

O/E:



HS I + II + 0

HR 76 SR

Patient comfortable and aware of plan

Plan:

1. Admit
2. NBM from 0200
3. To be consented by surgeons in AM
4. Needs pre-op ECG
5. Bloods sent – FBC, UE's, LFTs, Coag,

Dr. A Brown

Surgical FY1

Post-operative ward round

8/10/17 Ms. Adrienne Smith 1/6/87 MRN 6340897

19:00 WR Mr. B. Colon (Cons)

Day 0 post-op

Post IR resection – patient well

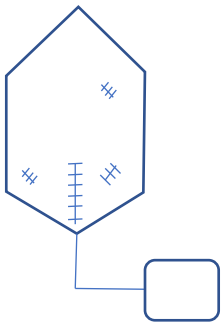
Comfortable in bed – using PCA

Has not yet sat out

Not tolerating oral fluids

O/E

Obs – BP 110/70 HR 96 Sat 100% 2L02



Generally tender – acceptable post-op

U/O 30 mls/hr

Impression: Patient well – try to encourage oral intake

Plan:

1. 8 hourly i.v. fluids
2. Sit out tomorrow AM
3. Start pentasa – written up
4. Continue PCA

Mr. A. Brown Surgical FY1

9/10/17 Ms. Adrienne Smith 1/6/87 MRN 6340897

14:00 WR Mr. B. Colon (Cons)

Day 1 post-op

Post IR resection

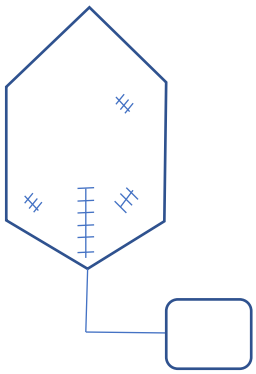
Patient comfortable – still using PCA

Tolerating oral fluids / soft foods

Reports passing flatus, BNO

O/E

Obs BP 120/80 HR 80 Sats 100% RA U/O 40 mls



Impression: Stable – doing well – consider taking down PCA tomorrow – will need to prescribe regular morphine based on PCA usage

Plan:

1. Physio
2. PCA down tomorrow
3. Stop IV fluids

Dr. A. Brown – Surgical FY1

Raw scoring for case 1

Patient records	Admission clerking	Correct
	Daily ward notes	Correct
	Operation note	Correct
	Admission bloods	Correct
	Drug chart	Correct
	Observations	Correct
History	Pain history	Correct
	Wound check	Correct
	Nausea and vomiting	Correct
	Respiratory Sx's	Correct
	Bowel habit	Correct
	Recent travel	Irrelevant
Examination	Respiratory	Correct
	Abdominal	Correct
	All others	Irrelevant
Bedside tests	CBG	Irrelevant
	ABG	Correct
	ECG	Correct

	Urine dip	Correct	
	MSU	Correct	
	All others	Irrelevant	
Request bloods	CRP	Correct	
	FBC	Correct	
	LFTs	Irrelevant	
	Coags	Irrelevant	
	U&E's	Irrelevant	
	Glucose	Irrelevant	
	Troponin	Irrelevant	
	All others	Irrelevant	
Further investigations	Chest radiograph	Correct	
	CT abdo/ Echo	Irrelevant	
	Colonoscopy/CT head/Bone scan/CT Chest	Incorrect	
Initial management	Oxygen	All options	Correct
	Nursing orders	NBM	Correct
		Repeat obs 1-2 hourly	Correct
		Repeat obs 8 hourly	Incorrect
		Stool chart	Irrelevant
		Urinary catheter	Correct

		NG tube insertion	Correct
		Fluid chart	Correct
	Medications	All analgesia	Incorrect
		Anti-emetics - cyclizine	Correct
		Metoclopramide	Incorrect
		Antibiotics – 1.2 g co-amoxiclav i.v	Correct
		All others	Incorrect
	Call colleague	N/A	
	Fluids	N Saline/Hartmann's 8 hrly	Correct
		All others	Incorrect
	Emergency	All options	Incorrect
Differentials	Cardiovascular	Acute MI	Correct
		All others	Irrelevant
	Respiratory	Hospital acquired pneumonia / pulmonary embolism	Correct
		All others	Irrelevant
	GI	Ileus	Correct
		All others	Irrelevant
	Neurological	All	Irrelevant

	Urology	UTI	Correct
		All others	Irrelevant
	Gynae	All	Incorrect
Further management	Take to theatre	Incorrect	
	Watch and wait	Incorrect	
	Start antibiotics	Irrelevant	
	Continue antibiotics	Correct	
Call colleague	Chest physio	Correct	

Appendix 3 – Case 3 (Mr. Barrett)

Info given:

- PC –previous history of biliary colic, USS – show stones, no CBD dilatation, LFTs N
- PMH – nil
- SH – smoker
- EtoH – 5 pints/night
- Occupation – medical student (QMUL)
- DH - nil

1/4

Admission Clerking

PC: Elective lap cholecystectomy

HPC: previous history of biliary colic 6/52 ago, USS – show stones, no CBD dilatation, LFTs N. Admission for elective lap cholecystectomy

2/4

PMH: Nil

Drugs: NKDA

SH: 3rd year medical student, non smoker, 20-25 units EtoH per week

4/4

O/E – Resp: N
CVS: N
Abdo: mild RUQ tenderness, abdomen soft
DRE not done
Plan: 1. Consented
2. Bloods – FBC, UE's, LFTs, Coag, G&S – sent
3. NBM from midnight

1/2
Daily Notes:

D1 – patient was nauseous – therefore kept in overnight, patient, complaining of vague abdominal pain, abdomen soft, wound ok

Plan – 1. Continue analgesia
2. Start light diet - aim home tomorrow as patient still nauseous
3. TWOC

2/2

D2 - Tolerating diet, passed urine, bowels not opened. Still complaining of vague abdominal pain 3-4/10, shoulder tip pain
Team – Plan: home later

ATSP = temp 37.6 HR 95

Signed:

1/2
Operation Notes

Lap cholecystectomy- catheter for duration of operation – full bladder. WHO – abx and TEDS/tinz / flo-trons through op

Findings: Mildly thickened GB with omental adhesions, Calot's triangle dissected, identified cystic duct and artery - clipped and divided, GB mobilised from liver bed with some ooze, removed in endo-bag, otherwise uneventful

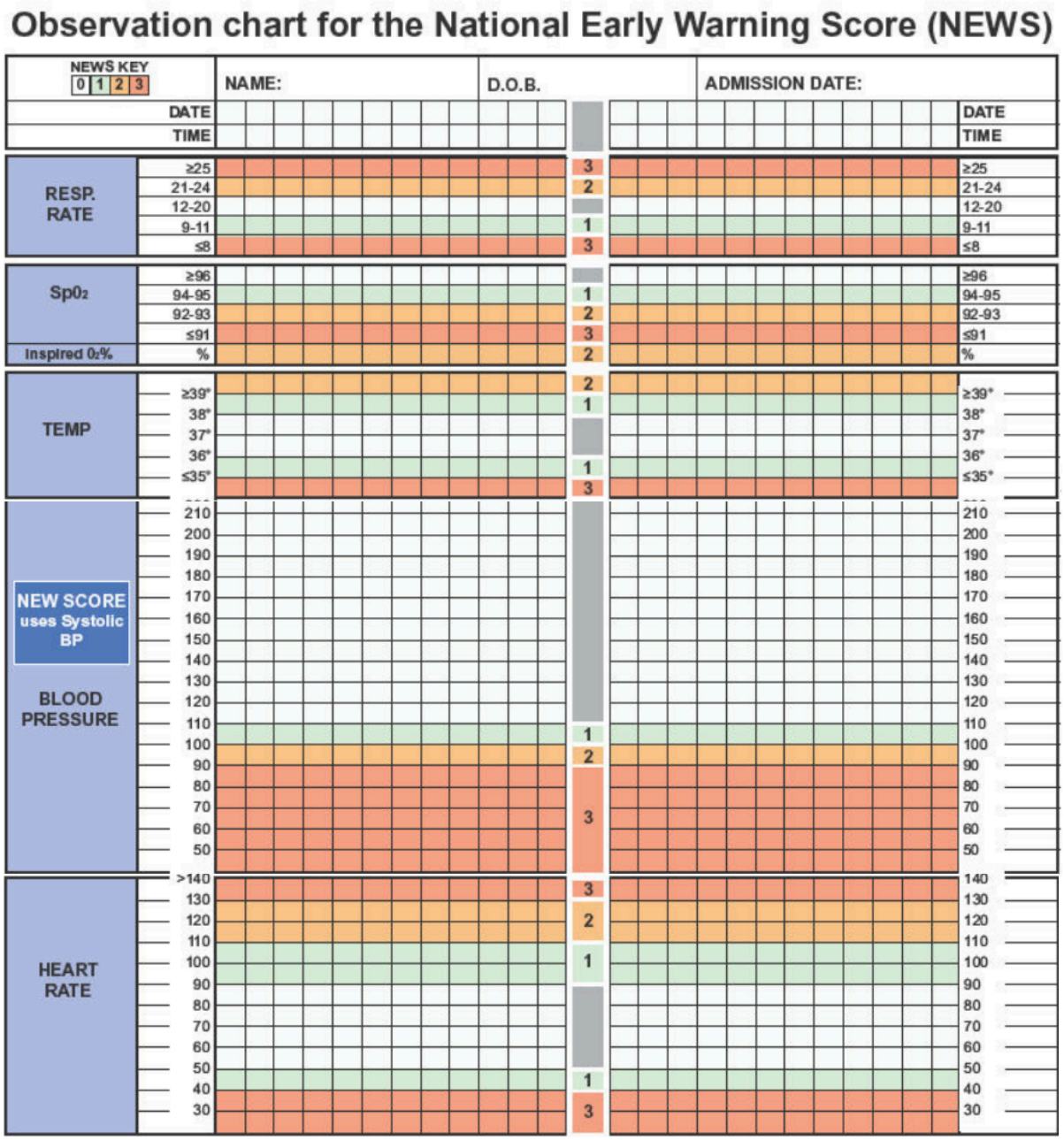
2/2
Operation Notes

Post-op instructions:

- 1. Eat and drink 2. TWOC then home

Signed:

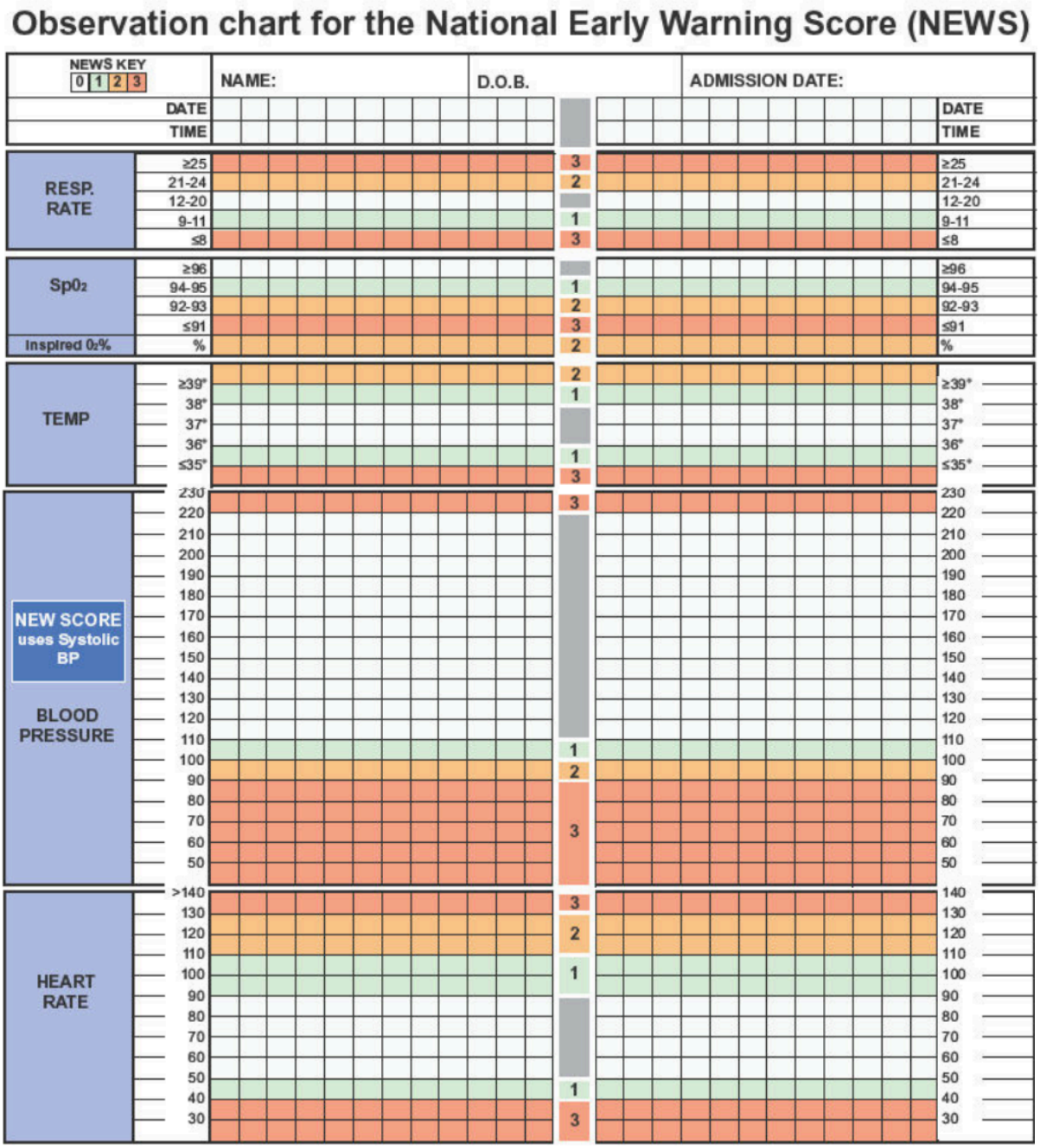
Obs 1:



Please note, the obs chart is split into three parts so that high resolution screen shots can be taken when zoomed in. Please add copies of this marker ● over image or list each trend.

Info given:

Obs 2:


Observation chart for the National Early Warning Score (NEWS)
NEWS KEY
0 1 2 3
NAME:
D.O.B.
ADMISSION DATE:
DATE
TIME
RESP. RATE
≥25
21-24
12-20
9-11
≤8
Sp0₂
≥96
94-95
92-93
≤91
Inspired 0₂%
%
TEMP
≥39°
38°
37°
36°
≤35°
NEW SCORE uses Systolic BP
BLOOD PRESSURE
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
HEART RATE
>140
130
120
110
100
90
80
70
60
50
40
30

Fluid Prescription Chart

Date	Fluid	Vol. (mL)	Additive name	Additive dose	Time (hours)	Route	Sig.
No fluids							

^information given: ‘8-hourly bags’

ABG: no needs

Parameter	Result	Units	Reference Interval

pH			7.35-7.45
pCO2		kPa	4.7-6.0
pO2		kPa	10.5-14.0
HCO3-		mmol/L	22-26
Lactate		X	X
Potassium		mmol/L	3.3-5.3
Base excess		mmol/L	22-30

Urine:

Parameter	Result
pH	
Glucose	
Leukocytes	+ve
Erythrocytes	+ve
Nitrites	-ve

FBC: pre-op N

Parameter	Result	Units	Reference Interval
RBC count		$x10^{12}/L$	3.8-5.8

WBC count	13	$X10^9/L$	4.5-11.0
Platelets	300	$x10^9/L$	140-400
Hb	130	g/L	115-165
Hematocrit	0.38		
MCV	82	fL	80-100
MCH	30	pg	27-32
MCHC	34	g/dL	32.0-36.0

LFTs: N

Parameter	Result	Units	Reference Interval
Total bilirubin	14	µmol/L	<20
ALT	28	IU/L	<33
AST	41	IU/L	<45
ALP	130	IU/L	30-130
Gamma GT	39	IU/L	<40
Albumin	49	g/L	35-50

U&Es: N

Parameter	Result	Units	Reference Interval
Creatinine	70	µmol/L	45-84
Na	136	mmol/L	133-146
K	4	mmol/L	3.5-5.3

CRP: Post-op

Parameter	Result	Units	Reference Interval
CRP	45	mg/L	<10

Parameter	Result	Units	Reference Interval
PT	12	seconds	11-14
INR	NA	-	4.5-11.0
APTT	26	seconds	24-37
TT	14	seconds	12.5-17.0
Fibrinogen	1.8	seconds	1.5-2.5

Please provide CXR - atelectasis right base & Normal CT Abdo

DD:

SIRS

Pneumonia

Bile leak

O/E – N reduced air entry bilaterally bases, Obs N

Plan:

Bloods, CXR, USS abdo – for suphrenic collection/ GB fossa collection, urine dip (trace blood,

WCC +) –ve all else

CXR – atelectasis,

USS – no collection

Plan:

Home as atelectasis

Appendix 4 - Demographics questionnaire

Age _____ years Sex M / F

Level of training Medical student ___yr Foundation 1 Foundation 2

CT SpR (St3-5) SpR (St 6+) Consultant

Have you worked on a general surgical firm before?

Have you taken part in virtual reality simulation before? Yes No

If so, how many times_____

Do you play computer games? Yes No | If so, how often? Daily, Weekly, Monthly, Yearly

Have you ever taken part in immersive simulation using a head-mounted display? Yes No

Appendix 5 – System Usability Scale

System Usability Scale

Instructions: Please for each one of the statements on the left mark on the right one box that best describes your agreement from 1 (strongly disagree) to 5 (strongly agree)

	Strongly Disagree				Strongly Agree
I think that I would like to use this simulator frequently.	1	2	3	4	5
I found the this simulation unnecessarily complex.	1	2	3	4	5
I thought the this simulator was easy to use.	1	2	3	4	5

I think that I would need the support of a technical person to be able to use this simulator.	1	2	3	4	5
I found the various functions in this simulator were well integrated.	1	2	3	4	5
I thought there was too much inconsistency in this simulator.	1	2	3	4	5
I would imagine that most people would learn to use this simulator very quickly.	1	2	3	4	5
I found the simulator very cumbersome to use.	1	2	3	4	5
I felt very confident using this simulator.	1	2	3	4	5
I needed to learn a lot of things before I could get going with this simulator.	1	2	3	4	5

Appendix 6 – Presence Questionnaire

Presence Questionnaire

1. Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place.

I had a sense of "being there" in this simulation

Not at all						Very much
1	2	3	4	5	6	7

2. To what extent were there times during the experience when the virtual environment was the reality for you?

There were times during the experience when the simulation was the reality for me...

At no time						Almost all the time
1	2	3	4	5	6	7

3. When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?

The simulation seems to me to be more like...

Images that I saw						Somewhere that I visited
1	2	3	4	5	6	7

4. During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere?

I had a stronger sense of...

Being elsewhere						Being in *the simulation*
1	2	3	4	5	6	7

5. Consider your memory of being in the virtual environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual

memory of the virtual environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.

I think of simulation as a place in a way similar to other places that I've been today...

Not at all						Very much so
1	2	3	4	5	6	7

w

6. During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

During the experience I often thought that I was really standing in the simulation

Not very often						Very much so
1	2	3	4	5	6	7

Appendix 7 – Enjoyment Questionnaire

Likert Enjoyment

I enjoyed this experience

Strongly disagree Disagree Neutral Agree Strongly agree

If given access to a library of such modules, I would you use them

Strongly disagree Disagree Neutral Agree Strongly agree

This is a useful way to learn how to interpret simple tests

Strongly disagree Disagree Neutral Agree Strongly agree

This method of teaching can be used to improve clinical decision-making

Strongly disagree Disagree Neutral Agree Strongly agree

What suggestions would you make for improving this method of teaching?

What did you enjoy most about this experience?

What did you enjoy most about this experience?

Appendix 8 – Face validity questionnaire

Face validity

	Strongly disagree						Strongly agree
1. The process of taking a history from the virtual patient is realistic	1	2	3	4	5	6	7
2. The process of examining the virtual patient is realistic	1	2	3	4	5	6	7
3. The process of obtaining different investigations and their results for the virtual patient is realistic	1	2	3	4	5	6	7
4. The process of managing the virtual patient is realistic	1	2	3	4	5	6	7
5. The simulated ward was a realistic representation of a real ward	1	2	3	4	5	6	7

Appendix 9 – Cybersickness Questionnaire

SIMULATOR SICKNESS QUESTIONNAIRE

Circle how much each symptom below is affecting you

right now

General Discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eye strain	None	Slight	Moderate	Severe
Difficulty focusing	None	Slight	Moderate	Severe
Salivation increasing	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty concentrating	None	Slight	Moderate	Severe
Fullness of head	None	Slight	Moderate	Severe
Blurred vision	None	Slight	Moderate	Severe
Dizziness with eyes open	None	Slight	Moderate	Severe
Dizziness with eyes closed	None	Slight	Moderate	Severe
Vertigo*	None	Slight	Moderate	Severe
Stomach awareness**	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to

vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of

nausea.

Appendix 10 – Self-assessment questionnaire

Category		N/A	1	2	3	4	5
History	Gathers appropriate information						
	Follows logical sequence						
Examination	Examines relevant system						
Investigation	Selectively orders appropriate diagnostic study						
Diagnosis	Makes appropriate diagnosis						

Management	Institutes appropriate treatment	Self-rating scale 1 – STRONGLY DISAGREE 2-DISAGREE 3 – NEUTRAL 4 – AGREE 5 STRONGLY AGREE
Overall	Progresses through case in logical sequence Demonstrates appropriate clinical decisions	

Appendix 11 - Raw data for presence questionnaire

Presence					
2D Med student	3D Med student	2D SHO	3D SHO	2D SpR	3D SpR
0.80952381	0.738095238	0.38095238	0.76190476	0.76190476	1
0.914285714	0.714285714	0.69047619	0.78571429	0.47619048	1
0.166666667	0.833333333	0.61904762	0.88095238	0.76190476	0.66666667
0.619047619	0.80952381	0.66666667	0.88095238	0.73809524	0.64285714
0.428571429	0.738095238	0.47619048	0.83333333	0.66666667	
0.952380952	0.904761905	0.9047619	0.61904762		
0.942857143	0.833333333		0.78571429		
0.833333333	0.952380952				

Appendix 12 - Usability raw data

SUS					
2D Med student	3D Med student	2D SHO	3D SHO	2D SpR	3D SpR
90	60	87.5	67.5	80	95
80	85	77.5	70	82.5	100
85	92.5	72.5	55	77.5	72.5
85	85	80	92.5	82.5	77.5
82.5	80	57.5	77.5	57.5	
85	75	75	75		
95	92.5		72.5		

92.5	95				

Appendix 13 - Cybersickness raw data

Cybersickness					
2D Med student	3D Med student	2D SHO	3D SHO	2D SpR	3D SpR
6 none	5 none	5 none	4 none	3 none	1 none
2 slight	2 slight	1 slight	3 slight	2 slight	3 slight
	1 moderate				

Appendix 14 - Face Validity Raw data

Face validity					
2D Med student	3D Med student	2D SHO	3D SHO	2D SpR	3D SpR
29	27	26	27	25	35
32	20	26	26	24	0
16	25	34	26	25	29
16	24	14	26	27	18
28	25	18	30	33	25
34	30	35	26		
33	29		20		
26	31				

Appendix 15 - Raw data presence

Presence

2D Med student	3D Med student	2D SHO	3D SHO	2D SpR	3D SpR
0.80952381	0.738095238	0.38095238	0.76190476	0.76190476	1
0.914285714	0.714285714	0.69047619	0.78571429	0.47619048	1
0.166666667	0.833333333	0.61904762	0.88095238	0.76190476	0.66666667
0.619047619	0.80952381	0.66666667	0.88095238	0.73809524	0.64285714
0.428571429	0.738095238	0.47619048	0.83333333	0.66666667	
0.952380952	0.904761905	0.9047619	0.61904762		
0.942857143	0.833333333		0.78571429		
0.833333333	0.952380952				

Appendix 16 - Script concordance test checklist

Table 2: Script Concordance Test item quality grid, adapted from Caire

Scenario	• Describes a challenging situation, even for experts	Yes / No
	• Describes an appropriate situation for examinees tested	Yes / No
	• The scenario is necessary in order to understand the question and to set the context	Yes / No
	The clinical presentation is typical	Yes / No

	The scenario is correctly written	Yes / No
Questions	Questions are developed following a key-feature approach	Yes / No
	In the experts' opinion, the options are relevant	Yes / No
	The same option is not found in two consecutive questions	Yes / No
	The new information (2nd column) makes it possible to test the link between the new information and the option (1st column) in the described context	Yes / No
	Likert scale anchors are clearly defined and unambiguous	Yes / No
	Questions are developed to spread the answers equally over all the values of the Likert scale	Yes / No

	Questions are developed to provide balance between low and high variability	Yes / No
Expert's panel	Number between 10 and 20	Yes / No
	The experts' panel includes experienced physicians whose presence in a jury is appropriate to the level of the examinees assessed	Yes / No
	Experts take the test individually, in exactly the same conditions as the examinees	Yes / No

Appendix 17 – Usability raw data (Contrast validity study)

	System usability		
	Medical student	FY2	CT2
1.	75	75	100
2.	85	75	75
3.	82.5	77.5	72.5

4.	75	92.5	77.5
5.	72.5	80	70
6.	80	80	57.5
7.	77.5	97.5	82.5
8.	60	100	75
9.	75	72.5	75
10.	67.5	92.5	80
11.	95	95	75
12.	82.5	82.5	72.5
13.	67.5	67.5	62.5
14.	80	67.5	100
15.	72.5	85	72.5
16.	65	90	70
17.		72.5	70
18.		72.5	45
19.		80	80
20.		47.5	
21.		87.5	

Appendix 18 – Self-scoring raw data (Contrast validity)

	Medical student	FY2	CT2

1.	3	2.5	4
2.	2	2	3.625
3.	3.125	3.125	3.625
4.	3.25	3.5	3
5.	2.75	3.25	3.875
6.	3.375	0	3.375
7.	3.5	2.875	4.625
8.	3	4.375	4.125
9.	3.625	3.375	3.875
10.	4.125	4.125	4
11.	3.75	3.875	4
12.	4	0	4
13.	3.25	3.625	3.5
14.	2.875	3.875	3.5
15.	2.375	4.375	4
16.	4	4	3.875
17.		3.875	2.875
18.		3.125	3.25
19.		3.75	3.25
20.		4	
21.		3.25	

Appendix 19 - Pre-teaching Questions

Question 1

Abdominal pain

Please answer the following questions about abdominal pain presentations – for each clinical vignette, choose the next appropriate investigation. All options can be used once, multiple times or not at all.

A. Abdominal radiograph
B. CT scan
C. CT angiogram
D. Erect chest radiograph
E. Exploratory laparoscopy
F. MR angiography
G. Serum amylase
H. Serum CRP
I. Supine chest radiograph

1. A 76 year old gentleman presents with a central abdominal pain that occurs 20 minutes after eating, he has noticed weight loss over the last few months. He suffers with intermittent claudication. On examination – his abdomen is soft and non-tender.

2. A 30 year old lady is admitted to hospital with acute onset severe abdominal pain and associated vomiting, she reports that the pain is radiating to her back, there is associated vomiting. Obs : HR - 120 regular BP 94/70mmHg. On examination - tenderness and guarding in the epigastrium.

3. An 84 year old gentleman with a history of knee osteoarthritis presents with acute onset severe abdominal pain and vomiting. Obs : HR 96 regular BP is 110/76mmHg. On examination – rigid abdomen with absent bowel sounds.

4. A 25 year old lady presents with sudden-onset right iliac fossa pain associated with nausea and vomiting. She had a similar admission one year ago which was diagnosed as torsion of an ovarian cyst. HR is 90, BP is 120/80mmHg. On examination she has tenderness and guarding in the right iliac fossa.

5. A 58 year old man presents with a one month history of dysphagia, initially to solids and now to liquids. He has a history of GORD for which he is on a PPI and has noticed significant weight loss. Endoscopy shows a cessile mass in the lower oesophagus

Question 2

Hypersensitivity

Please answer the following questions about hypersensitivity presentations – match the clinical vignette with one of the options. All options can be used once, multiple times or not at all.

A. Type I
B. Type II
C. Type III
D. Type IV
E. Type V
F. Not a hypersensitivity reaction

1. A 25 year old develops wheezing, glossitis and a rash following a bee sting.

2. A 35 year old lady presents with a short history of haematuria, urine dipstick shows microscopic haematuria. She reports having time off work with an upper respiratory tract infection a few weeks ago.

3. A 30 year old lady presents with an 8 week history of diarrhoea, weight loss and feeling warm all the time. On examination she has a diffuse symmetrical midline goitre.

4. A 40 year old general surgeon presents with diffuse rashes over both hands, he believes it is related to his glove use.

5. A 38 year old multi-parous woman gives birth to a stillborn baby. The mother is known to be Rhesus-negative and the baby was to found to be Rhesus positive.

Appendix 20 - Script concordance test 1

A 30 year old lady presents with a 24 hour history of lower abdominal pain, temp 37.6 C

If you were thinking of:	And then you find	The diagnosis becomes				
Appendicitis	Her last menstrual period was 2/52 ago	-2	-1	0	+1	+2
Ruptured ectopic pregnancy	Polycystic ovarian syndrome	-2	-1	0	+1	+2
UTI	Surgery for ureteric reflux as a child	-2	-1	0	+1	+2

Investigative

60 year old man 5 days post emergency open right hemi-colectomy spikes a temperature of 38.5

If you were thinking of ordering a:	And then you find	The investigation becomes
Chest radiograph	His saturations are 99% on room air	-2 -1 0 +1 +2
Urine dipstick	He has a catheter in situ	-2 -1 0 +1 +2
CT abdo/pelvis	He has gas under the right diaphragm on the chest radiograph	-2 -1 0 +1 +2

A 30 year old lady presents with a 24 hour history of lower abdominal pain, tender and guarding in the right iliac fossa

If you were thinking of ordering a:	And the result is:	The diagnosis of x becomes:
Urine dipstick	Leucocyte +ve, erythrocyte – ve, nitrite -ve	UTI: -2 -1 0 +1 +2 Appendicitis -2 -1 0 +1 +2
USS abdomen	Free fluid in the pelvis	UTI: -2 -1 0 +1 +2 Appendicitis -2 -1 0 +1 +2

Amylase	120	Pancreatitis					
		-2	-1	0	+1	+2	

Appendix 21 - Script concordance test 2

Please complete the following questions, indicating if the statement in the third column is:

-2 much less likely

-1 less likely

0 – has no effect

+1 – more likely

+2 – much more likely

60 year old man 3 days post emergency open right hemi-colectomy spikes a temperature of 38.5C, high NG output

If you were thinking of:	And then you find	The diagnosis becomes				
Hospital acquired pneumonia	The patient doesn't report a cough	-2	-1	0	+1	+2
Urosepsis	He has a catheter in situ draining clear urine	-2	-1	0	+1	+2
Anastomotic leak	He has not opened his bowel or passed flatus	-2	-1	0	+1	+2

A 30 year old lady presents with a 24 hour history of lower abdominal pain, temp 37.6 C

If you were thinking of:	And then you find	The diagnosis becomes				
Appendicitis	Her last menstrual period was 2/52 ago	-2	-1	0	+1	+2
Ruptured ectopic pregnancy	Polycystic ovarian syndrome	-2	-1	0	+1	+2
UTI	Surgery for ureteric reflux as a child	-2	-1	0	+1	+2

Investigative

60 year old man 5 days post emergency open right hemi-colectomy spikes a temperature of 38.5

If you were thinking of ordering a:	And then you find	The investigation becomes				
Chest radiograph	His saturations are 99% on room air	-2	-1	0	+1	+2
Urine dipstick	He has a catheter in situ	-2	-1	0	+1	+2
CT abdo/pelvis	He has gas under the right diaphragm on the chest radiograph	-2	-1	0	+1	+2

A 30 year old lady presents with a 24 hour history of lower abdominal pain, tender and guarding in the right iliac fossa

If you were thinking of ordering a:	And the result is:	The diagnosis of x becomes:
Urine dipstick	Leucocyte +ve, erythrocyte – ve, nitrite -ve	UTI: -2 -1 0 +1 +2 Appendicitis -2 -1 0 +1 +2
USS abdomen	Free fluid in the pelvis	UTI: -2 -1 0 +1 +2 Appendicitis -2 -1 0 +1 +2
Amylase	120	Pancreatitis -2 -1 0 +1 +2

Appendix 9

Appendix 22 - Clinical reasoning problems

You are called to see a patient who is 4 days post laparoscopic left hemi-colectomy as he has spike a temperature of 38.6C

What are your top two differentials?

1.

2.

List the three most discriminatory tests that you could request?

1.

2.

3.

You are called to see a 26 year old sexually active lady who presents with severe lower abdominal pain and a temperature of 38.5C.

What are you top two differentials?

1.

2.

What are the three most discriminatory tests that you could perform?

1.

2.

3.

Appendix 23 – Pre-question raw data (Chapter 6 study)

Raw pre-question

Pre-questions VR	Pre-questions iCBD
6	2
3	6
5	6
7	6
6	8
4	4
5	8
7	6

6	9
3	6
6	6
5	2
6	5
4	3
5	5
6	6
4	1
4	6
6	3
8	6
10	6
3	3
5	4
7	1
6	3
6	

Appendix 12

Appendix 13

Info given:

- Admission clerking –
- 48 hrs after admission – currently under observation, not on antibiotics, receiving i.v. fluids
- Awaiting gynae opinion from another hospital
- Bloods:
- WCC 12.5 CRP 30

1/4

Admission Clerking

PC: Abdominal pain

HPC: 25F admitted with lower abdominal pain, worse on the right, nauseated, vomited once, mild pyrexia 37.5 C, LMP – 2/52 ago. Previous episode of abdo pain around at the time of her menses without vomiting, pain worse this time. BO loose for last 24 hrs
Sexual history: sexually active, nulliparous

2/4

PMH: nil else

DH: NKDA

SH: Environmental consultant, EtoH – few glasses of wine per week, non-smoker.

3/4

O/E – Resp N
CVS N

Abdo – generalized tenderness lower abdomen, right and left side equally tender, no rebound, no guarding, Rovsing's negative

Plan:

1/2
Daily Notes:

D1 – same as before, vomited once more, bowels remain loose, still awaiting gynae opinion.
Pelvic USS reported – poor views of the pelvis due to empty bladder, no obvious adnexal masses, appendix not visualised, some free fluid in pelvis ?physiological

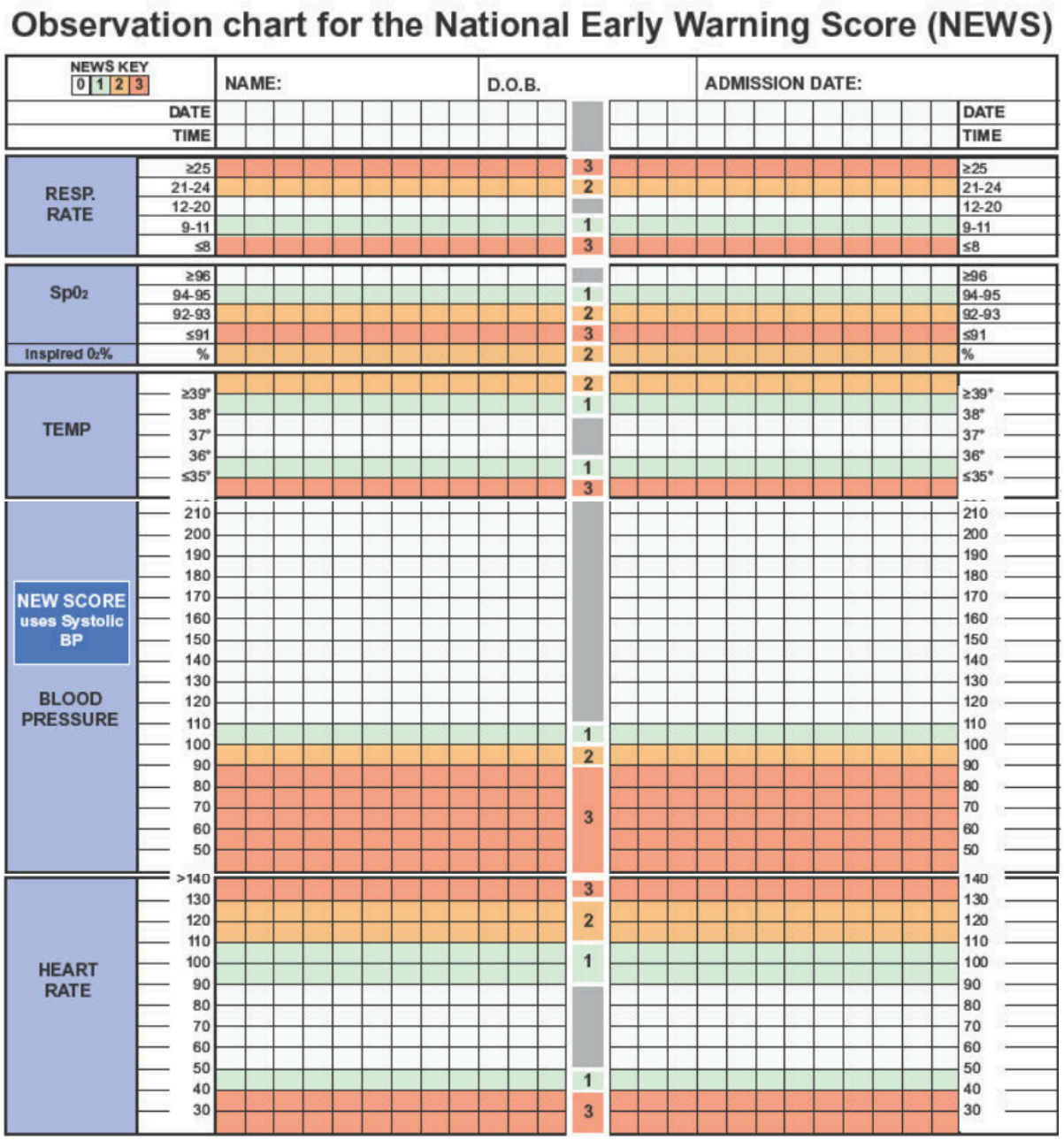
Gynae team review – hx noted – speculum / bimanual examination revealed no cervical excitation / no discharge.
Pelvic USS report noted – no evidence of gynae pathology as cause of symptoms.
Plan: Continue care under general surgical team.

D2 –pain worse, now spiking temp 38.6, HR 110

2/2

Signed:

Obs 1:



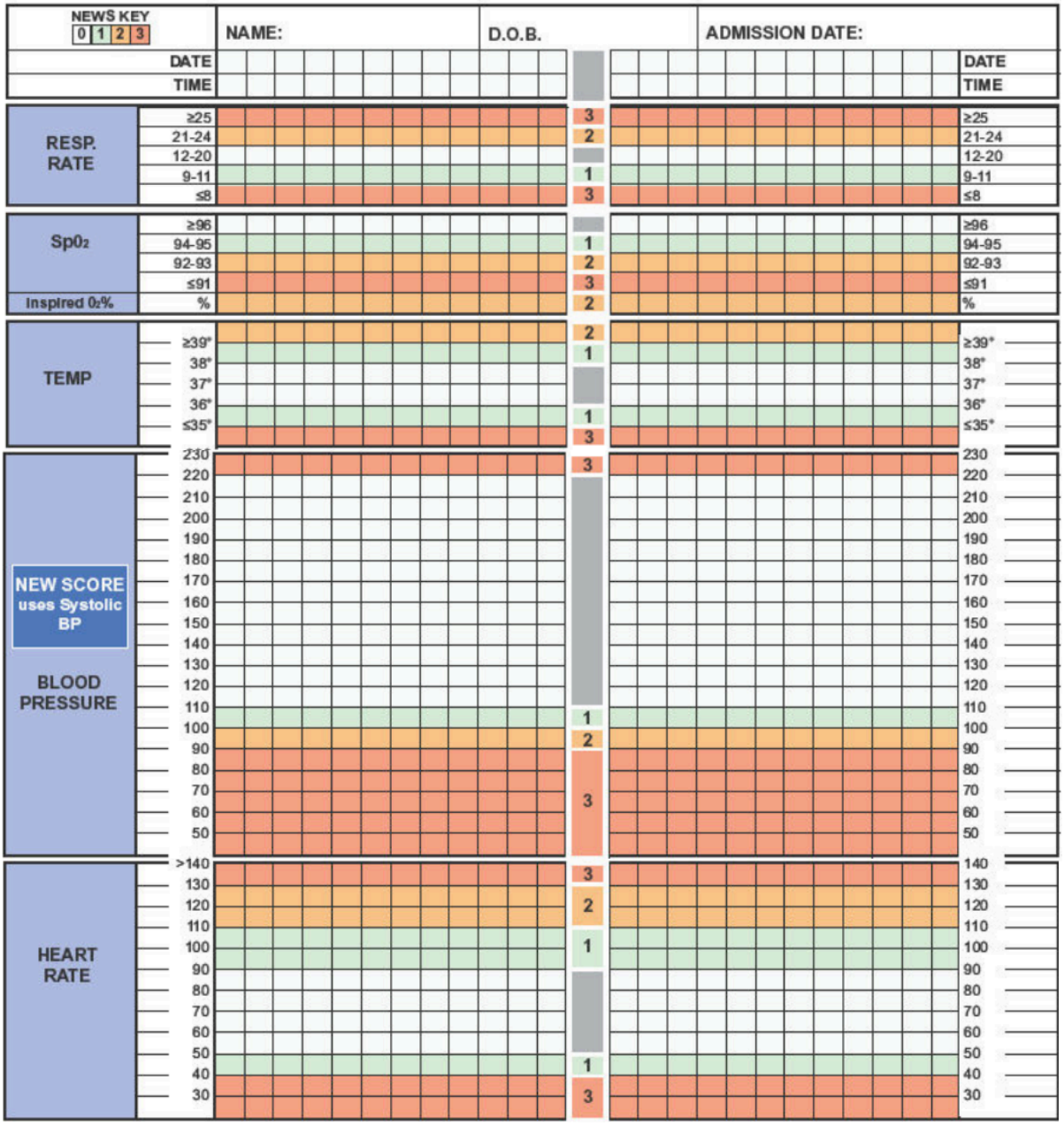
Please note, the obs chart is split into three parts so that high resolution screen shots can be taken when zoomed in. Please add copies of this marker ● over image or list each trend.

Info given:

- 90
- RR – 10 constant
- BP – N

Obs 2:

Observation chart for the National Early Warning Score (NEWS)



Fluid Prescription Chart

Date	Fluid	Vol. (mL)	Additive name	Additive dose	Time (hours)	Route	Sig.

^information given: ‘8-hourly bags’

ABG:

Parameter	Result	Units	Reference Interval
pH			7.35-7.45
pCO2		kPa	4.7-6.0
pO2		kPa	10.5-14.0

HCO3-		mmol/L	22-26
Lactate		X	X
Potassium		mmol/L	3.3-5.3
Base excess		mmol/L	22-30

Parameter	Result
pH	
Glucose	-ve
Leukocytes	+ve
Erythrocytes	-ve
Nitrites	-ve

O/A

Parameter	Result	Units	Reference Interval
RBC count		$x10^{12}/L$	3.8-5.8
WBC count	11.5	$X10^{9}/L$	4.5-11.0
Platelets	380	$x10^{9}/L$	140-400
Hb	110	g/L	115-165

Hematocrit	0.35		
MCV	85	fL	80-100
MCH	30	pg	27-32
MCHC	33	g/dL	32.0-36.0

LFTS:

Parameter	Result	Units	Reference Interval
Total bilirubin	15	µmol/L	<20
ALT	20	IU/L	<33
AST	24	IU/L	<45
ALP	86	IU/L	30-130
Gamma GT	31	IU/L	<40
Albumin	38	g/L	35-50

U&Es:

Parameter	Result	Units	Reference Interval
Creatinine	70	µmol/L	45-84
Na	135	mmol/L	133-146

K	5	mmol/L	3.5-5.3

CRP:

Parameter	Result	Units	Reference Interval
CRP	30	mg/L	<10

Coag screen:

Parameter	Result	Units	Reference Interval
PT	12	seconds	11-14
INR	N/A	-	4.5-11.0
APTT	24	seconds	24-37
TT	13	seconds	12.5-17.0
Fibrinogen	1.6	seconds	1.5-2.5

Bloods now D2

Parameter	Result	Units	Reference Interval
WBC count	16	$X10^9/L$	4.5-11.0

Platelets	390	$x10^9/L$	140-400
Hb	110	g/L	115-165
Hematocrit	0.4	L/L	3.8-5.8
MCV	85	fL	80-100
MCH	31	pg	27-32
MCHC	33	g/dL	32.0-36.0

Parameter	**Result**	**Units**	**Reference Interval**
CRP	120	mg/L	<10

Please provide values for: CBG

CBG =

Please provide ECG and CXR

CT report

Reported by:

Now – temp with pain.

Hx – should have urinary symptoms – frequency with minimal voiding

Rigors

O/E Suprapubic – local peritonism

Ix: Urine dip – exclude pyelonephritis

Potentially CT

DD – pelvic abscess with secondary

?start antibiotics????

References