



Getting Behind the Headset: Exploring Clinical Reasoning in Virtual Reality Simulation.

Dr Lauren Medwell

Note that some graphs/tables/images have been removed in order to comply with copyright restrictions.

MSc In Medical Education, 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION



1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	
Date:	07/12/2025

Table of Contents

Acknowledgements	10
Abstract	11
Introduction	13
Literature Review	15
Section One - Clinical Reasoning	15
<i>Defining Clinical Reasoning</i>	15
<i>Theories Regarding Clinical Reasoning</i>	17
The Process of Reasoning	17
Knowledge Structures	18
<i>Illness Script Theory</i>	19
<i>Pattern Recognition Model</i>	20
<i>Summary of Knowledge Structures</i>	20
Dual Process Theories	21
What About Context?	22
State Versus Trait	22
<i>Approach of this Research to the Concept of Clinical Reasoning</i>	23
Section Two - Teaching and Assessing CR	25
Section Three - Clinical Reasoning in Simulation	28
Section Four - Virtual Reality	29
<i>Virtual Reality Simulation</i>	30
<i>Virtual Reality and Clinical Reasoning</i>	30
VR Simulation and CR	31
Summarising the Literature and Considering the Gap	34
Research Aims	35
Methodology	35
Overall Approach	35
Research Methodology	36
Research Methods	37
<i>Baseline Questionnaire</i>	37
<i>Mixed Methods</i>	37
Sampling	39
Exclusion Criteria	39
Recruitment	40

Research Process	40
<i>The Simulation</i>	41
<i>Pre Brief and Practice</i>	42
<i>Researched Simulation</i>	42
Data Collection.....	44
<i>Aspect One: Actions and Outcomes</i>	44
Model Answer.	45
<i>Aspect Two: Sequential Flow</i>	47
<i>Aspect Three: Think Aloud</i>	47
Analytical Strategy.....	49
<i>Baseline Questionnaire</i>	49
<i>Integrated Analysis</i>	49
<i>Aspect One: Actions and Outcomes</i>	49
<i>Aspect Two: Sequential Flow</i>	50
<i>Aspect Three: Think Aloud</i>	51
Quality and Rigour.....	52
<i>Reflexivity Statement</i>	53
Ethical Considerations	54
Risks	54
Consequences.....	54
Potential Perceived Conflict of Interest.....	54
Data Storage and Protection.	55
Changes and Reflections	55
Results.....	57
Participant Description.....	57
<i>Demographics</i>	57
<i>Prior Experience</i>	58
Overview of Results.....	59
<i>Aim 1: To explore how clinical reasoning can be demonstrated in VR simulation.</i>	59
<i>Aim 2: To investigate how clinicians of different expertise levels approach clinical reasoning in VR simulation.</i>	59
The Overall Flow and Key Actions.....	61
<i>Analysis of Overall Flow and Structure</i>	62
<i>Use of A-E Structure</i>	62
Theme One: Not Just Theories - How Data Demonstrated Clinical Reasoning Related to Theories and Models	64

<i>Hypothetico-deductive Reasoning</i>	64
Hypothesis Testing.	65
<i>Pattern Recognition</i>	66
Pattern Recognition in Interpretating Investigations.	66
<i>Use of Mental Schemas</i>	68
Conditional Reasoning as a Mental Schema.	68
More Complex Conditional Reasoning Schema	68
<i>Use of Illness Scripts</i>	70
Less Developed Illness Script.	71
Unverified Assertions.	71
<i>Dual Process</i>	72
<i>Problem Representation</i>	73
Theme Two: Indicators of CR Errors - Common Cognitive Pitfalls	74
<i>Outcomes and Actions</i>	74
Diagnosis.....	74
Omissions in Investigative and Management Actions	75
Completion of Sepsis Six.	76
<i>Errors in Cognitive Process</i>	77
Errors in Inferences	77
Errors in Interpretation.	77
<i>Use of Guidelines and Choice of Management</i>	78
Scoring Systems as a Decision Aid.	78
Theme Three: The Context - How VR Simulation may Impact the CR Process	81
<i>Technology Specific Issues</i>	81
<i>Fidelity</i>	81
Communication Fidelity.....	81
Examination Fidelity.	82
Investigation Fidelity.....	83
Discussion	85
Aim 1: To explore how clinical reasoning can be demonstrated in VR simulation.	85
<i>Theme One: Not Just Theories - How Data Demonstrated Clinical Reasoning Related to Theories and Models</i>	86
Hypothetico-deductive Reasoning.	86
Pattern Recognition.	86
Conditional Reasoning Schema.	87
Illness Scripts.	88

Dual Process.....	89
<i>Theme Two: Indicators of CR Errors - Common Cognitive Pitfalls.....</i>	<i>90</i>
<i>Theme Three: The Context - How VR Simulation may Impact the CR Process.....</i>	<i>93</i>
Implications: Where Might VR Simulation Fit within CR Education?	96
Clinical Reasoning Concepts.	96
History and Physical Examination.	97
Choosing and Interpreting Clinical Tests.	97
Problem Identification and Management.	98
Shared Decision Making.	99
<i>Concluding Aim One.....</i>	<i>99</i>
Aim 2: To investigate how clinicians of different expertise levels approach clinical reasoning in VR simulation.....	100
<i>Sequential Flow.....</i>	<i>101</i>
Feature One: Differences in Overall Consultation Flow.	101
Feature Two: Use of Explicit Structures.	102
<i>The Think Aloud.....</i>	<i>102</i>
Feature Three: Pattern Recognition in Diagnosis.....	103
Feature Four: Differences in Interpreting Investigations.	103
<i>Outcomes and Actions.....</i>	<i>104</i>
Feature Six: Diagnostic Accuracy – Identification of Sepsis.	104
Feature Seven: Differences in Average Percentage of NICE Recommended Actions Completed.	104
Implications for Assessment of CR.....	107
Limitations	107
Conclusion	108
References.....	108
Appendices.....	121

TABLE of FIGURES and TABLES	Page Number
FIGURE 1: Researcher understanding of interplay between theories and models of CR	23
FIGURE 2: Domains of CR education	25
FIGURE 3: Advantages and disadvantages of VR	29
FIGURE 4: Six characteristics of mixed methods	37
FIGURE 5: Data triangulation	38
FIGURE 6: Research process	40
FIGURE 7: Meta Quest headset	41
FIGURE 8: VR simulation sepsis scenario image	43
FIGURE 9: Sepsis Six pre 2024 version	46
FIGURE 10: Analysis Steps	51
FIGURE 11: Features of expertise along a continuum	60+100
FIGURE 12: Overall simulation flow and key actions by participant	61
FIGURE 13: Time to mention key differentials by participant	64
FIGURE 14: Domains of CR that VR simulation is most suited to	96
TABLES	
TABLE 1: Percentage of correct decision and estimated mortality in a simulated trauma scenario by expertise level in Colonna et al (2022)	33
TABLE 2: Main activity categories	50
TABLE 3: Summary of participant prior experience	58
TABLE 4: Summary of features mentioned in handover	70
TABLE 5: Percentage of model answer actions completed based on NICE guidelines	75
TABLE 6: Completion of sepsis six by participant	76
TABLE 7: Calculation of scoring systems and antibiotic decisions by participant	79

APPENDICES	Page Number
APPENDIX ONE: Baseline Questionnaire	123
APPENDIX TWO: Poster Advertisement	127
APPENDIX THREE: Participant Information Sheet	128
APPENDIX FOUR: Consent Form	133
APPENDIX FIVE: Model Answer Checklist	135
APPENDIX SIX: Categorisation of Activities	136
APPENDIX SEVEN: Codebook	137
APPENDIX EIGHT: Example Research Notes	141
APPENDIX NINE: Example Reflexivity Log During Analysis	142
APPENDIX TEN: Ethical Approval Letter	144
APPENDIX ELEVEN: Local Approval Letter	145
APPENDIX TWELVE: Risks and Mitigations	146
APPENDIX THIRTEEN: Recruitment Flow Sheet	147

Acknowledgements

This master's degree has been a great experience, which would not have been possible without the help of some incredible people both on my academic journey but also those who kept the other plates spinning!

My first and biggest thanks goes to my participants. Thank you for giving up your precious time and being willing to have a go. I am also grateful to Goggleminds^{Ltd} for allowing use of their virtual reality simulation and being on hand for technical support.

My thanks to my supervisor, Dr Sally Shiels for her guidance and knowledge of virtual reality. I am incredibly grateful to all the tutors on the master's programme who turned my interest in medical education into a lifelong passion.

I would also like to thank Dr Chris Jacobs for his mentorship which sparked my interest in this field of research, his advice has been invaluable. I am grateful for the whole team at the education academy at Great Western Hospital for supporting this project.

Thanks to my course mates who I learnt so much from. Thanks for the laughter and mutual support at the times we were all questioning our decisions.

To my family and friends, thanks for believing in me. You have been the steady presence holding things together so I could embark on this project. Particular thanks to my fiancé for taking on the mundane tasks of life and being there in my moments of doubt. To my mum, I can never repay your unwavering support and quite literally building a house around me as I worked!

Abstract

Background: Clinical Reasoning (CR) is central to health professionals' competence. CR is poorly defined but includes gathering, synthesising and processing information to diagnose and treat patients. Erroneous reasoning has been associated with significant mortality and morbidity worldwide. Developing CR requires opportunity for explicit repeated practice. Simulation is known to be an excellent tool for teaching CR but is a limited, costly resource. Virtual Reality (VR) simulations are the latest tool, with initial research suggesting a role in CR education. However, no research has explored the CR process in depth within VR simulation, making it hard to consider its best use case.

Aims:

- 1.) To explore how clinical reasoning can be demonstrated in VR simulation.
- 2.) To investigate how clinicians of different expertise levels approach clinical reasoning in VR simulation.

Methodology: This study used a convergent mixed methods design to explore CR in depth within a single VR simulation regarding sepsis. To consider the impact of expertise, participants were from three groupings: medical students, early career clinicians, and experienced clinicians. Participants were asked to think aloud whilst completing a simulation in VR. Three data sources were triangulated and analysed in an integrated way: sequential flow, outcomes and actions, and the think aloud data.

Results: Ten participants took part. The results showed clear evidence of CR demonstration. All participants considered sepsis apart from the two most junior, however, the average percentage of expected actions completed showed an inverse

relationship with expertise. Seven features differentiating expertise levels were found as well as three themes regarding how CR can be demonstrated in VR simulation:

- 1.) Not just theories: how data demonstrated CR related to theories and models;
- 2.) Indicators of CR errors: common cognitive pitfalls;
- 3.) The context: how VR simulation may impact the CR process.

Discussion: This study concludes that CR can be demonstrated within VR including identification of erroneous reasoning. Results indicated great potential as a teaching tool, especially in CR domains related to considering investigations, problem identification and management. This learning could best be leveraged with inclusion of expert debrief to dissect the why behind actions. The context of VR simulation did impact on the CR process, particularly the scaffolding of questions and drop-down menus. This makes it a less useful tool for teaching CR domains such as communication and shared decision making. There were some features seen with increasing expertise. However, the data was divergent with different data sources highlighting greatest success in different groups. This suggests a need for caution in using VR simulation to assess CR, as different assessment metrics may produce contrasting results.

Conclusion: Overall, this study adds to the body of evidence that VR simulation can be an excellent tool in CR education, especially paired with expert debrief. Nevertheless, using this tool within assessment should be considered with caution to make the most of VR simulations advantages, whilst acknowledging its limitations.

Introduction

Clinical Reasoning (CR) is an essential skill to ensure safe, effective healthcare. Poor CR leads to diagnostic error which has been linked to significant mortality and morbidity worldwide (WHO, 2025). In the US, it was estimated that 40,000 hospital deaths annually resulted from diagnostic errors (Tehrani et al., 2013). Autopsy studies report discrepancy between presumed diagnosis and postmortem findings of up to 39% (Aalten et al., 2006). Even if diagnosis is correct, errors in management reasoning can lead to incorrect action. McGlynn et al. (2003) found that, on average, adults received only 55% of recommended care. There has long been interest in CR, how it develops and how we can assess and teach these skills.

Medicine is both a science and an art (Panda, 2006) with CR falling into the grey zone that ties these together. Whilst objective data and eventual diagnosis seem inherently scientific, doctors do not deal with data, they deal with humans. CR seeks to decode people and, as I often tell my students, our patients do not read the textbooks, so may not present in a typical way. I once saw a patient where every objective measure suggested one conclusion, but the context, their views and my gut feeling said otherwise. Teaching CR involves explaining that there are as many exceptions as rules, and the value of that hard to learn skill, intuition. This leads us to a paradox; CR is essential to clinical competence but how do we teach such a nebulous thing? This difficulty has led to CR teaching being haphazard or not being taught explicitly at all.

A recent UK consensus report (Cooper et al., 2021) identified five domains of CR education that need embedding throughout curriculums. Whilst students require opportunity for repeated, contextualised practice; clinical pressures and growing student numbers makes this challenging. Simulation helps develop CR and improves outcomes, especially with repeated sessions (Theobald et al., 2021). Virtual Reality (VR) is emerging as a new tool for delivering simulation opportunities with greater resource efficiency, adaptability and repeatability (Pottle, 2019). Whilst research around VR is growing, most studies focus on student experiences or self-reported metrics rather than in-depth investigations with consideration of theory (Curran et al., 2022). Initial studies suggest potential for VR simulation having a role in CR education, yet minimal data exists, especially regarding its best use.

Researching CR, a poorly defined phenomenon, in the field of VR simulation, presents challenges. VR simulation is being used now, with its development being driven commercially, so it is important to pause to understand the underlying pedagogical implications. I approached this research as an educator and clinician who has used VR personally. The purpose of this in-depth study is to determine how CR is demonstrated in a VR simulation and to compare the approach used by novices and experts. I hope to add to the evidence base by getting behind the headset and linking the findings to theory.

Literature Review

This literature review explores aspects of both CR and VR necessary to contextualise the study. First, it considers the definition of CR and related theories. This sets out the theoretical framework underpinning this research. The literature review then focusses on how CR is taught and assessed, particularly in relation to simulation. It defines VR, and its use in simulation before finally focussing on the current research regarding CR within VR simulation.

Section One - Clinical Reasoning

This section considers the definition of CR; related theories and the working concept of CR used in this research. CR is widely recognised as essential to health professionals' roles and development of competence (Jay et al., 2024). Despite research interest in CR, the literature is fragmented with no consensus on the underlying process, what reasoning entails and even how to define CR.

Defining Clinical Reasoning

CR lacks a single universally accepted definition. Young et al. (2019) identified 110 different terms used for CR with no joint consensus on categorising these terms amongst experts. In addition, definitions differ across healthcare professions. Simmons (2010) explored CR in nursing, emphasising that reasoning should not be conflated with clinical decision making or judgement which focus on reasoning outputs. Within

physical therapy, a concept analysis defined CR as a cognitive and psychomotor process that is contextual, involving multiple perspectives with the goal being a biopsychosocial approach to patient management (Huhn et al., 2019). This holistic interpretation of CR contrasts with the traditional physician-orientated view which focusses on a clinician's internal mental process to reach a differential diagnosis (Gordon et al., 2022). It could be argued that interprofessional variety reflects different professional roles. Historically, diagnostics was undertaken by physicians, possibly explaining the value placed on diagnosis as an outcome.

Even amongst physicians there is little consensus on the definition of CR. Working definitions of CR vary from anything resulting in diagnosis or therapy to purely the internal cognitive process (Young, Thomas, et al., 2018). Additionally, management CR has been argued to be different to diagnostic reasoning, as the skills needed to create a treatment plan are complex with multiple correct outcomes (D. Cook et al., 2018). Young et al. (2018) posited that definition variation may be related to how broadly individuals consider the boundary of what to include in the CR definition. Although, a broad definition which defines almost everything as CR could undermine its practical application (Gruppen, 2016). Cooper et al. (2021) similarly argue that the fragmented nature of the literature has hindered implementing CR research into practical education.

It is clear that there is no one accepted definition and that the complexity obfuscates the nature of CR. Overall, I agree with Huhn et al. (2019) that attempting to strictly define such a complex construct seems limiting. As our understanding is still developing, a working concept of CR provides greater flexibility. Therefore, rather than selecting one definition, this literature review considers underlying theories underpinning CR to explain the working concept used within this research.

Theories Regarding Clinical Reasoning

There are numerous interconnected theories and models underpinning our understanding of CR. Yazdani & Hoseini Abardeh (2019) categorised them into three areas: theories on the **process of reasoning**, theories on **knowledge structure** and **compilation theories** such as dual process. This section explores the connections between these theories and why one in isolation is inadequate.

The Process of Reasoning. Yazdani & Hoseini Abardeh (2019) concluded that only the hypothetico-deductive model considered the overall process of CR. The model was proposed by Elstein et al. (1978) and suggests that clinicians generate a small number of hypotheses early from initial data and then test these. A separate study with medical students concluded a similar model (Neufeld et al., 1981). These studies used a technique called Think Aloud (TA). This is a method in which participants are asked to verbalise their internal monologue whilst completing a task (Noushad et al., 2024). It is ideal for considering CR and has been used extensively in research.

The hypothetico-deductive model has been criticised for failing to explain distinctions between expert and novices. Patel & Groen (1986) highlighted this was not reflective of expertise in other fields. Outside medicine, forward reasoning, a process where causal rules are applied to known facts, was used by experts; whereas backward reasoning, generating and then testing hypotheses, was used by novices. Barrows et al. (1982) suggested that experts may make faster and better initial hypothesis but failed to explain how this may occur. Further criticism of hypothetico-deductive reasoning comes from a lack of content specificity, where accurate diagnosis in one case was a poor predictor in another, despite the same reasoning process being demonstrated (Elstein et al., 1990; Higgs, 1992). The hypothetico-deductive models' simplicity is both a strength and a weakness. It describes a systematic reasoning process, and is a useful strategy for students and within complex cases (Elstein, 1994). However, the model's inability to discriminate expertise or link to strategy success, limits its ability to explain CR.

Knowledge Structures. Research in the 1980s shifted to understanding how clinicians structured knowledge and the role of memory (Norman, 2005b). Studies regarding expertise in other fields had suggested the importance of memory, not necessarily related to direct recall. In chess, expert players did not have better general memory than novices but instead had better recall of meaningful information organised in patterns (Chase & Simon, 1973). Earlier experimentation demonstrated that humans recall the essence of information in a shorter format rather than literally (Bartlett, 1932/1954). This led to the concept of a schema, as basic representations people use to remember

gists or scripts, a term for a more specific type of knowledge structure (Custers, 2015). In a medical context, Charlin et al. (2000) describes schemas as goal directed knowledge structures, similar to if/then flow charts.

Illness Script Theory. Feltovich & Barrows (1984) proposed that illness scripts are a form of mental schema consisting of three parts: enabling conditions (modifiable/ non modifiable risk factors), the fault (underlying pathophysiology) and the consequences (signs and symptoms). Schmidt et al. (1990) proposed experts have greater experiential knowledge and therefore more developed and efficiently structured illness scripts. Interestingly, several studies showed that underlying pathophysiology (the fault) was given less emphasis with increasing expertise (Custers et al., 1996, 1998). In summary, this theory suggests that clinicians draw from organised knowledge structures (scripts) to guide CR during consultations.

Illness script theory had implications for medical education. It suggested focus should be on teaching knowledge representation rather than a generic reasoning process. This means that concept linking is as essential as giving the building blocks of knowledge. Mandin et al. (1997) suggested students should be taught schemas as an organised way to approach problems, as each presentation may require different strategies rather than one process. Despite this, empirical evidence is lacking for the effectiveness of using illness script theory within teaching and experimental studies showed mixed results (Delavari et al., 2020; van Schaik et al., 2005; Wu et al., 2025).

Pattern Recognition Model. Another knowledge structure related model is the pattern recognition model in which clinicians compare the features of a patient presentation with a stored bank of disease patterns (Barrows & Feltovich, 1987). Case et al. (1988) suggested that pattern recognition is often used by expert physicians. However, evidence for this only exists in certain fields, particularly those involving visual reasoning (Norman et al., 2007). Furthermore, Elstein (2009) questioned pattern recognition as it cannot explain the complete CR process and there is little understanding for when it is used, preferable or even correct.

Summary of Knowledge Structures. Theories of CR related to the structuring of knowledge clearly have some merit in their practical useage in teaching and assessment. Nevertheless, they fail to fully explain the CR process and lack empirical evidence outside of specific fields.

Dual Process Theories. These theories originate in older philosophical ideas, but were comprehensively described in relation to medicine by Wason & Evans (1974). Reasoning processes were split into two different modes: system one, typically fast and intuitive and system two, slower and more analytical. Pattern recognition has been suggested as the CR process underlying system one, with hypothetico-deductive reasoning being the process for system two (Pelaccia et al., 2011). Eva (2005) suggests that both processes form part of CR for expert and novice clinicians. There are several criticisms of dual process theory. Firstly, the numerous theories have different conceptions, descriptions and language making it challenging to define or investigate (Evans & Stanovich, 2013). Keren & Schul (2009) further argue that dual process theories lack conceptual clarity and empirical evidence.

Dual process theories have been linked to medical error (system one and system two errors), especially as error is frequently related to deficits in clinicians' cognitive processes (Graber et al., 2005). System one, or intuitive thinking, was perceived to be more error prone, with improvement strategies focused on reducing cognitive bias and encouraging slower thinking (Croskerry, 2009; Nendaz & Perrier, 2012). Norman & Eva (2010) and Norman et al. (2024) criticised this, identifying several studies that conclude intuitive diagnosis are often correct (Barrows et al., 1982; Sherbino et al., 2012). Moreover, experimental evidence does not show that encouraging slower thinking improves diagnostic accuracy (Ilgen et al., 2013; Norman et al., 2014). In visual reasoning fields, such as radiology, there is evidence that slower, analytic thinking can increase normal variants being identified as pathological, especially by

less experienced clinicians (Monteiro et al., 2024). Several empirical studies have identified increased diagnostic accuracy if both strategies are used concurrently (Eva et al., 2007; Norman et al., 1999). In summary, dual process theory provides a good platform for considering CR processes and has led to positive consideration concerning error reduction.

What About Context? The theories discussed so far focus on the internal mental processes and exclude how context may impact CR. Durning et al. (2013) highlighted the importance of context and presented an alternative view of CR utilising situativity theory. This theory considers that all learning is situated in experience and environments (Durning & Artino, 2011). For CR, this means that patient and environmental factors need consideration. Situativity theories have been criticised for lacking consideration of cognitive processes (Yakhlef, 2010). However, they may better reflect the complexity of true clinical practice. This has implications for using VR simulation within education. If the context impacts CR, the potential impact of the educational environment provided within VR needs thought.

State Versus Trait. One fundamental question that remains is whether CR is a state, which is context-dependent and variable; or a trait, with an individual having generic reasoning ability (Elstein et al., 1978). If it is a trait, tests such as the script concordance test, which assess the closeness of the CR process to that of an expert, should tell us an individuals' reasoning ability (Charlin, Roy, et al., 2000). Alternatively, as suggested by the lack of content specificity (Elstein et al., 1990), is CR a state which is context

dependent? Durning et al. (2013) would advocate for this view, however, the complexity of this approach may limit its practical application. Overall, CR can be considered a state, albeit that generic aspects of the internal cognitive process may be applied across multiple situations. This consideration led to an adaptation of my initial research question. I had planned to investigate CR abilities; however, the evidence does not support that an individual has a specific measurable CR ability. Instead, my research has considered context dependent demonstrations of CR.

Approach of this Research to the Concept of Clinical Reasoning

There is no consensus regarding the definition, boundaries and underlying theoretical understanding of CR. Nevertheless, as with any complex construct there is value in considering it through these different lenses. I have summarised my understanding of the interplay between the various theories in figure 1 below.

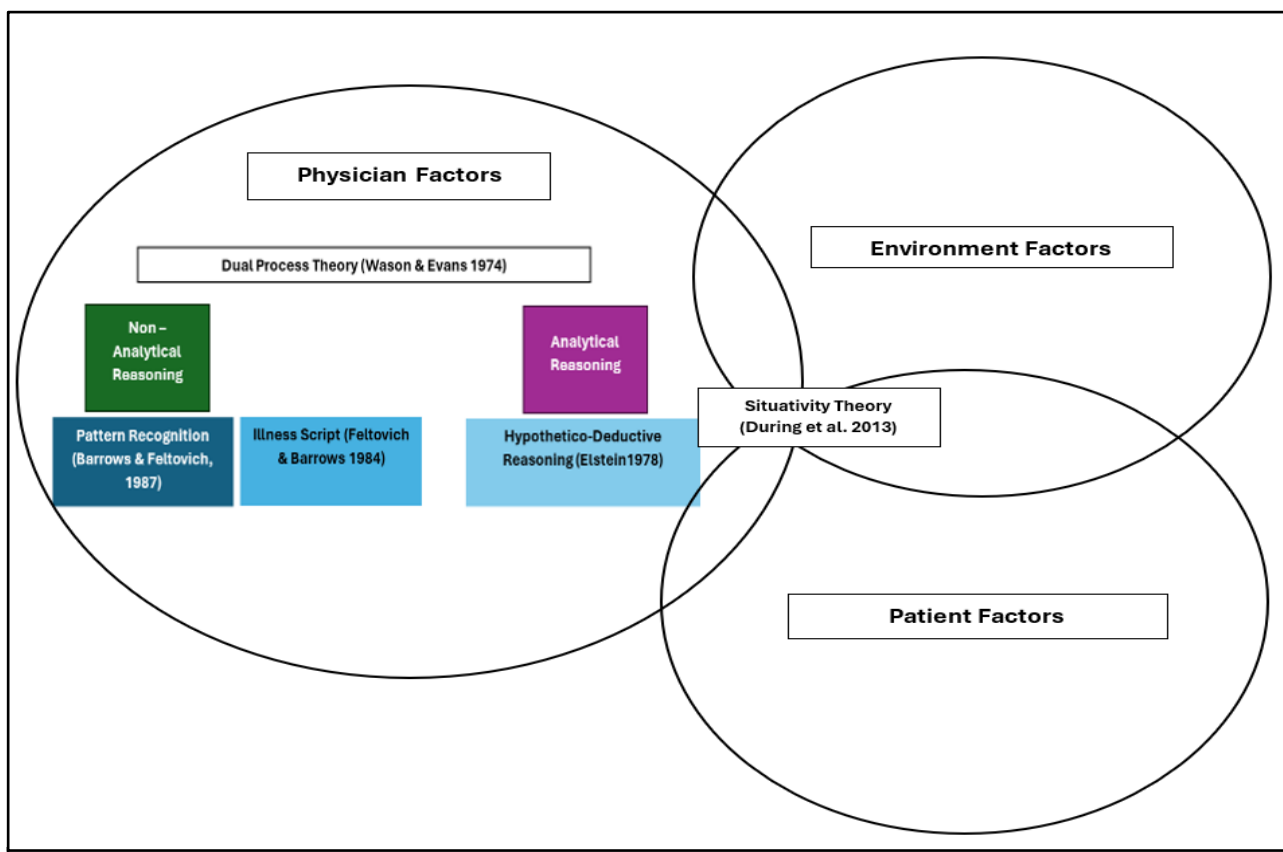


FIGURE 1: Researcher understanding of interplay between theories and models of CR

This research takes a broad view of CR including context, as reducing reasoning to its constituent parts risks losing the essence of the phenomenon. The working definition of CR I have adopted is most closely summed up in this concept definition: *“a skill, process, or outcome wherein clinicians observe, collect, and interpret data to diagnose and treat patients. Clinical reasoning entails both conscious and unconscious cognitive operations interacting with contextual factors.”* (Daniel et al., 2019, p.902)

Section Two - Teaching and Assessing CR

Despite recognition of CR's importance, there is no agreement on how to teach or assess it. A consensus statement from the UK Clinical Reasoning in Medical Education Group highlighted the importance of integrating CR throughout the curriculum within five domains (figure 2) (Cooper et al., 2021). Cooper & Frain (2022) found that current CR education is haphazard and that learning within clinical placements depends on case exposure and feedback opportunity. Classroom settings present clinical cases to students pre-structured, giving limited opportunities to practice problem representation. This provides a way to produce a mental representation of key case features to summarise and reframe the information. Opportunities to practice with numerous cases in different contexts is essential (Cooper & Frain, 2022).

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Cooper, N., Bartlett, M., Gay, S., Hammond, A., Lillicrap, M., Matthan, J., & Singh, M. (2021). Consensus statement on the content of clinical reasoning curricula in undergraduate medical education. *Medical Teacher*, 43(2), 152–159. <https://doi.org/10.1080/0142159X.2020.1842343>

There are innumerable opinions regarding the best way to teach CR. These range from classroom to clinical based education strategies, often grounded in case-based approaches (Wagner et al., 2024). Several papers give explicit guidance for educators on how to enhance CR teaching (Cutrer et al., 2013; Rencic, 2011; Young, Dory, et al., 2018). Currently, there is no clear evidence that teaching CR strategies alone can improve reasoning (Cooper et al., 2021; Sherbino et al., 2014). Whereas strategies focusing on building knowledge structures are more effective (Schmidt & Mamede, 2015). Gruppen (2016) suggests the idea of teaching CR as a generalisable skill is flawed, as the foundational research highlights no consistency in CR between cases and contexts.

Assessment of CR is seen, by some, as the holy grail of medical education (Penne & Anderson, 2020). Paper based assessments offer consistency but take a partial view as they breakdown CR into constituent parts (Daniel et al., 2019). Additionally, written assessments focus on general rather than case specific CR, with the script concordance test being commonly used (Brentnall et al., 2022; Charlin, Roy, et al., 2000). Simulated cases are felt to better align to practice whilst still maintaining an ability to control context (Daniel et al., 2019). Despite approaching CR as a whole process, simulation-based assessments are resource intensive and have low correlation with other assessment metrics (Daniel et al., 2019). This reduces their suitability for summative assessments. Schuwirth (2009) also noted the issue of the intermediate effect, a known phenomenon where, in simulation-based assessments, intermediately experienced individuals outscore experts (Schmidt et al., 1988). This reduces construct

validity as the assessment does not discriminate between expertise levels (Schuwirth, 2009). Ward based assessments of CR are the most authentic but are limited by suitable cases and time pressure (Daniel et al., 2019). In practice, a combination of these assessment methods is needed. As CR is a complex construct, one assessment is unlikely to address all aspects whilst being a valid and reliable metric.

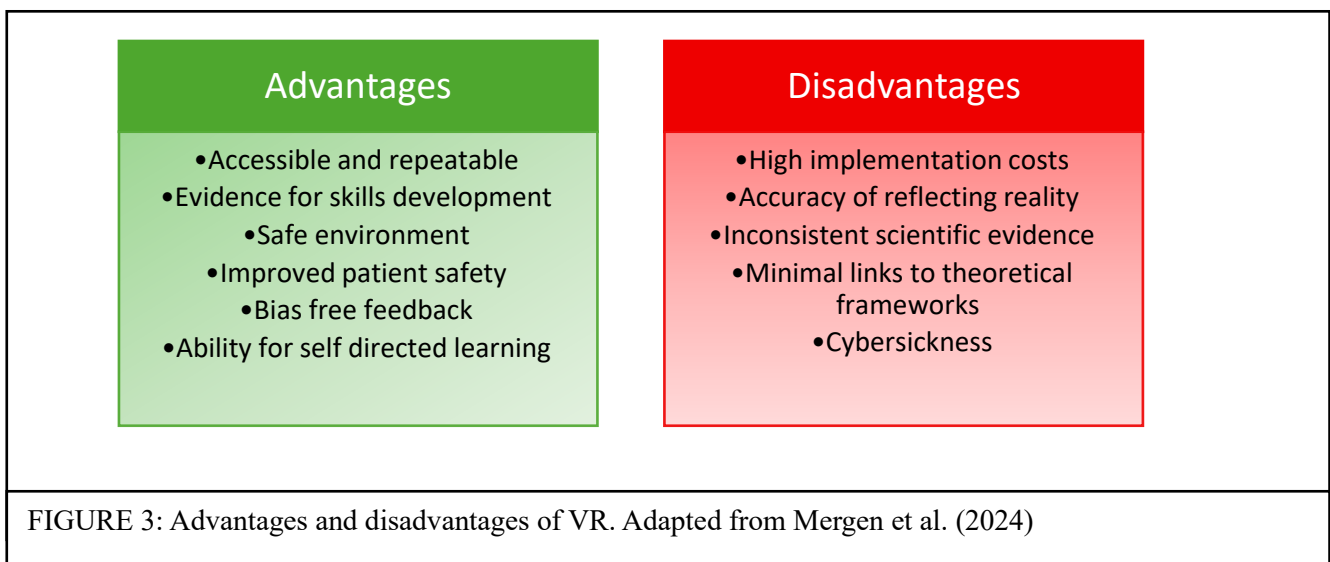
Section Three - Clinical Reasoning in Simulation

Simulation-based teaching seeks to mimic the clinical environment to provide immersive, experiential learning opportunities (Al-Elq, 2010). This section specifically considers high-fidelity simulation. Fidelity is defined as the degree to which a simulation reflects reality and has importance for learning (Tun et al., 2015). High-fidelity simulation has been used to help develop CR skills with evidence of effectiveness, especially with repeat simulations (Macauley et al., 2017; Theobald et al., 2021). A particular benefit of simulation in teaching CR is the whole case approach. Nendaz et al. (2000) demonstrated that when medics were presented a patient's main complaint with opportunity to ask questions, rather than a summarised case vignette, diagnostic accuracy dropped across all expertise levels. Presenting the case from an initial presenting complaint, like in simulation, challenges learners to work through the full CR process.

Despite the positives, simulation has challenges. Firstly, the evidence does not show any statistically significant gains to CR with single simulation exposure (Theobald et al., 2021). High-fidelity simulation is also resource intense with significant set up, space and staffing costs (Senvisky et al., 2024). Additionally, there is an argument that simulation learning comes from the post debrief, rather than the activity itself. Indeed, the debrief is widely considered the most critical section of a simulation (Issenberg et al., 2005; McGaghie et al., 2010). In summary, there is mixed evidence, but overall simulation can provide another useful tool for students to develop CR.

Section Four - Virtual Reality

Whilst simulation is a recognised way to teach CR, opportunities are limited, partly due to high resource requirements (Almousa et al., 2021). Extended reality, including virtual, mixed, and augmented reality is a novel teaching modality within medical education, and may help overcome some of the barriers (Curran et al., 2022). Extended reality is an umbrella term to describe the differing digital immersion levels (Milgram & Kishino, 1994). Virtual reality (VR) is entirely computer generated with complete immersion (Palmas & Klinker, 2020). VR technologies are being increasingly implemented within medical education (Mergen et al., 2024). Despite advantages, the scientific evidence base remains inconsistent and rarely links to theoretical frameworks. Research has focussed on learner reactions and knowledge acquisition, rather than more complex areas of medical practice, like CR (Barteit et al., 2021; Curran et al., 2022; Tang et al., 2022). Additionally, automatically generated feedback from VR applications raises concerns since it may be incorrect or potentiate bias (Combs & Combs, 2019). Indeed, Al-Saud et al. (2017) found that participants performed less well when no expert feedback was provided alongside the automatic system. Advantages and disadvantages of VR are summarised in figure 3.



Virtual Reality Simulation

The blurred lines of what is termed VR within literature makes it challenging to find relevant research. Virtual simulation has been used to describe everything from fully immersive systems using a Head Mounted Display (HMD), to partially immersed screen-based systems, to game base learning (Foronda, 2021). Gutiérrez et al. (2007) reported that fully immersive VR using a HMD enhances learning versus screen-based tools, suggesting that they should not be considered identical. Foronda (2021, para.2) stated that *“most simulationists would now interpret virtual reality simulation to involve a 360° immersion offered through a head mounted display that may or may not come with a haptic device”*, this is the definition I have adopted for my research.

Virtual Reality and Clinical Reasoning

Studies have explored using VR, virtual simulation and virtual patients to teach CR. Whilst these do not meet my definition for VR simulation, they are related and have relevance. Cook & Triola (2009) suggested that virtual patients are ideal for teaching CR because they allow for repeated practice with case variability. A metanalysis within nursing found significant improvement in CR in virtual simulation versus control (Sim et al., 2022). The majority used objective measures increasing authenticity. However, only two included studies were true immersive simulation. Interestingly, they found immersive VR to be less effective than screen-based, postulating this may be due to high cognitive load (Sim et al., 2022). Cognitive load theory considers that our cognitive network can only process a certain number of

information chunks concurrently (Sweller, 1988). However, we can overcome this by accessing organised chunked information to process tasks (van Merriënboer & Sweller, 2005). Evidence suggests virtual patient tools can increase medical students' CR skills within specific areas such as data gathering, rather than general measures such as problem solving (Plackett et al., 2022). Although these studies do not address my definition of VR simulation, they do provide evidence of successfully teaching CR with virtual tools.

VR Simulation and CR. Searching the literature identified studies in multiple healthcare fields regarding CR in immersive VR simulation using HMD.

Multiple studies have used validated self-reported CR tests as their outcomes measure. Self-reported scores ask participants to report their own abilities and are often tested before and after an intervention. All found improvements in self-reported scores either addressing overall CR (Havola et al., 2021; Jeong et al., 2022; Yang & Oh, 2022) or aspects of CR such as decision making (Hartstein et al., 2022; Kim et al., 2023; Lee et al., 2023), problem solving (Yang & Oh, 2022), and confidence at diagnosis/management (Jayasundera et al., 2022). Havola et al. (2021) defended the use of self-reported scores because they correlated with those given by the automated simulation scoring system. However, Keicher et al. (2024) reported self-assessments showed poor correlation with objective measures. Although findings were suggestive of value in CR education, the self-reported scoring measures only provide weak evidence for VR's use.

Only one study took a purely qualitative approach. Pardue et al. (2023) explored development of clinical judgement after a VR immersive simulation experience by retrospectively exploring problem-solving approaches. They found narrative evidence that VR simulation can assist in the development of CR, and that students were able to consider the reasoning behind decisions. Although the retrospective nature may have impacted their findings.

Other studies considered more objective measures of reasoning. Written or practical tests were used to compare VR simulation to current teaching methods. Two showed equivalent improvements in CR between VR and control (Donathan et al., 2023; Hartstein et al., 2022). One showed significantly better outcomes with VR (Baidoo & Adu, 2025) and one showed inferior outcomes with VR (Bonnin et al., 2023). The study demonstrating inferiority used a VR 360 video and did not have truly interactive elements (Bonnin et al., 2023). These studies generally suggest a positive role for using VR to teach CR and used objective, not self-reported metrics. Nevertheless, these tests considered CR as a general ability, which is not backed up by the known variety in CR success between cases.

Some studies focused on outcomes of CR and considered accuracy in diagnosis or triage. Equivalence in triage accuracy when taught using VR versus control was shown in both nursing (Ferrandini Price et al., 2018; Lee et al., 2023) and paramedic studies (Mills et al., 2020). In medicine, VR significantly outperformed standard teaching for medical students' ability to recognise crucial findings including

respiratory distress (Zackoff et al., 2020). Abulfaraj et al. (2021) reported no significant difference in time to perform critical seizure management actions between VR intervention and control. Outcomes are an essential component of CR, as they show final output, but are only a partial measure of the reasoning process.

Three studies considered a sequence or set of actions. Rochlen et al. (2023) utilised sequential behaviour analysis to investigate clinical decision making. They identified sequences of behaviour common to successful and unsuccessful simulation outcomes. Keicher et al. (2024) looked at VR emergency scenarios with junior doctors. General stabilisation tasks were well executed but only 65.6% of required actions were completed. Researchers proposed that extending reasoning to complex case management presented greater challenge for juniors, but they did not have a comparative group. This makes concluding that results were due to inexperience conjecture. A VR study regarding advanced trauma life support similarly considered the percentage of correct decisions as time to decision, and estimated percentage mortality (Colonna et al., 2022). They compared four levels of medics and were able to distinguish novices and experts in several metrics, summarised in table 1.

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Colonna, A., Robbins, R., Stefanucci, J., Creem-Regeh, S., Patterson, P., Engel, B., Fischer, A., & Nirula, R. (2022). Trauma bay virtual reality-A game changer for ATLS instruction and assessment. *The Journal of Trauma and Acute Care Surgery*, 93(3). <https://doi.org/10.1097/TA.0000000000003569>

Expert discrimination adds to the validity of using VR as a CR educational tool. Nevertheless, considering actions does not explain the reasoning behind decisions. It is too simplistic to classify clinical decisions as strictly right or wrong as things are rarely this simple in clinical practice.

Summarising the Literature and Considering the Gap

In summary, CR is central to clinical competence and patient safety. However, the only consensus regarding CR is its importance. Several theoretical frameworks and models of CR have been suggested and evaluated with varying degrees of success. They do not oppose each other and view aspects of CR through differing lenses. The lack of agreement on where the boundaries of CR fall, fragments the literature but also exemplifies the complexity of the phenomenon. Notwithstanding more academic arguments, there is a clear need for CR education and assessment; several tools have been employed to this effect. Traditional simulation is effective in CR but is resource intensive. VR simulation is the newest tool in the arsenal, with promising potential for increasing accessibility. Whilst a growing body of evidence suggests a role for VR in CR education, much of this is based on self-reported metrics, post-tests, or pseudo measures of reasoning such as diagnostic outcomes. To date, no study has taken a comprehensive review of the CR process in VR simulation, which combines actions and outcomes with the why behind these. The why can only be elucidated through qualitative discovery. By pairing this with objective outcome measures and sequential flow within a mixed methods study, it should be possible to explore a more detailed picture.

Research Aims

- 1.) To explore how clinical reasoning can be demonstrated in virtual reality simulation.
- 2.) To investigate how clinicians of different expertise levels approach clinical reasoning in virtual reality simulation.

Methodology

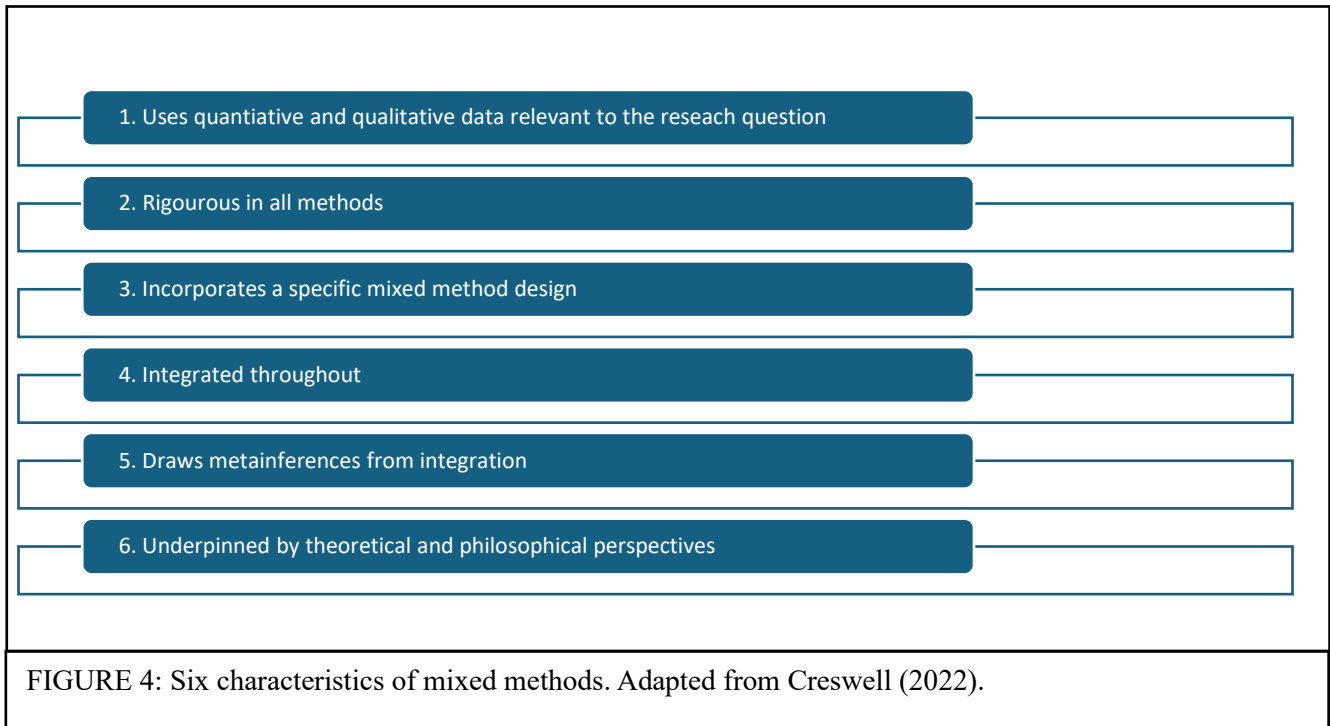
Overall Approach

To address the research aims this study took a mixed methods approach, triangulating three different aspects of data to gain an in-depth appreciation of CR in VR simulation. I approached this research through a pragmatist paradigm, focused on the practical applicability of findings, whilst accepting from a philosophical perspective that there can be both singular and multiple realities (Creswell & Plano Clark, 2018). From an ontological perspective, pragmatists accept an objective reality external to the researcher, but that this is encountered through human experiences constructed in social environments (Kaushik & Walsh, 2019; Morgan, 2013; Tashakkori & Teddlie, 1998). Pragmatists doubt the nature of reality can conclusively be determined (Pansiri, 2005) and so the axiology centres on the belief that value is determined by practical effects. Powell (2001, p.884) stated that *“To a pragmatist, the mandate of science is not to find truth or reality, the existence of which are perpetually in dispute, but to facilitate human problem-solving.”* Pragmatist epistemology does not equate knowledge to reality, as knowledge is based on unique individual experiences (Kaushik & Walsh, 2019). Knowledge is constructed through experience in order to

better engage in the world (Goldkuhl, 2012). Therefore, the best way to approach the nature of knowing, is in whichever way produces the most useable knowledge. Pragmatism best reflected my own ontological and epistemological views and underpinned my study design, with a goal to increase useable knowledge within medical education. Furthermore, it offered flexibility in approach to a complex phenomenon like CR, especially when considering it in VR – a novel way humans can experience reality.

Research Methodology

This study utilised a mixed methods approach due to the value of triangulating data. CR is multifaceted involving many aspects between an initial problem and eventual outcome. By integrating multiple methods, it provided the opportunity not only to understand the process, but also the reasoning behind actions. Durning et al. (2013) suggested that mixed methods was the ideal way to investigate CR as it is a complex non-linear phenomenon. Mixed methods are not just the combining of methods, but its own methodological orientation with shared techniques (Creswell & Plano Clark, 2018). Creswell (2022) summarises six key characteristics within this methodology which I aimed to address (figure 4).



Research Methods

Baseline Questionnaire

To contextualise the data, a questionnaire was conducted using a Microsoft Forms survey (appendix 1). It collected basic demographics, clinical experience, University or speciality, as well as approximate standard simulation and VR simulation experience.

Mixed Methods

This research employed a convergent mixed methods design (Creswell & Plano Clark, 2018). It triangulated three aspects (figure 5) with data collected simultaneously within one VR simulation before results were merged and interpreted together. This was considered most appropriate for comprehensively understanding CR within a

single VR simulation. Additionally, it best leveraged the potential advantages of a mixed methodology for triangulating data, and where pairing methods can help counteract the weaknesses of each (Wasti et al., 2022).

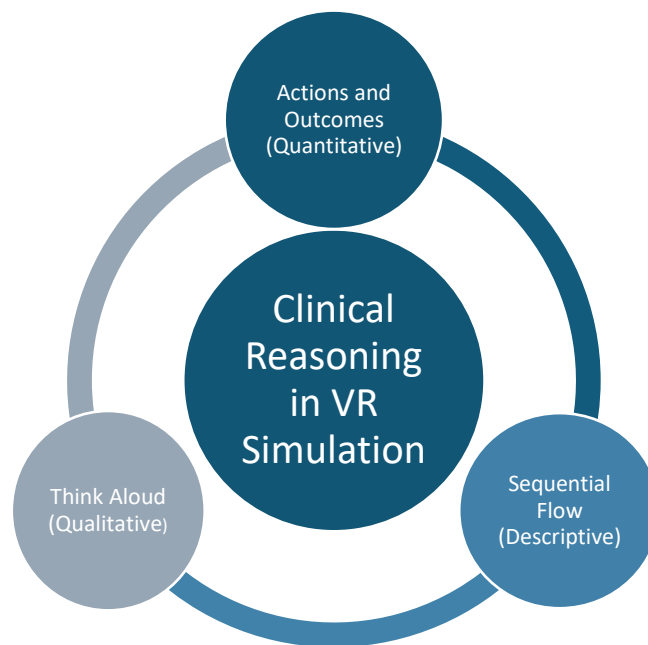


FIGURE 5: Data triangulation

To give appropriate context, the participant sampling, overall research process, and VR simulation are described below:

Sampling

Purposive criterion sampling was used for this study. This sampling method was chosen due to its benefit when needing to identify well-defined groups of participants (Memon et al., 2025), such as the expertise groupings needed to answer the second research question. The study took place at Great Western Hospital, a medium sized hospital with a large teaching academy. Whilst CR is relevant for all healthcare professionals, this study focuses on CR within medical professionals as this is my area of work.

Group one consisted of medical students. Recruitment was open to all years from: the University of Bristol, the University of Oxford and King's College London.

Group two consisted of qualified doctors and was planned to be solely clinicians at registrar or consultant level. Due to low recruitment numbers this criterion was expanded to any qualified doctor.

This led to **three analysis groupings**: medical students (any year), early career (pre specialisation within 5 years of qualification), and experienced clinicians (registrars and consultants).

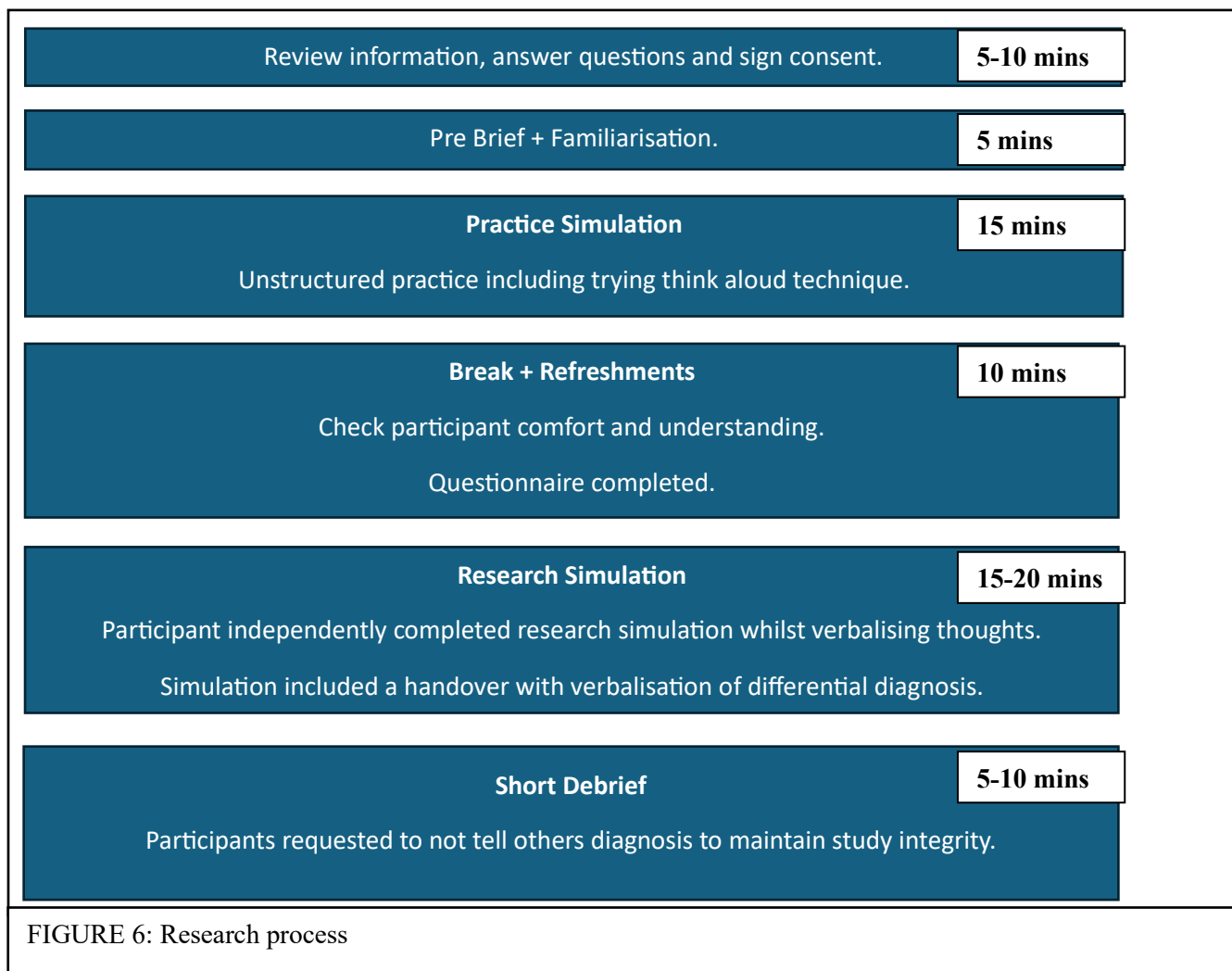
Exclusion Criteria. Previous participation in the specific Goggleminds^{Ltd} VR simulation being researched. As CR was the phenomenon of interest prior knowledge of the diagnosis would affect study results and integrity.

Recruitment

Physical and electronic poster advertisements (appendix 2) were distributed. Email distribution was sent by the medical education administration team to all undergraduate students, clinical teaching fellows, and the medical education consultant committee. Additionally, the researcher and administration team advertised in person. Snowball recruitment was also used. Once participants expressed interest, they were sent an introduction, participant information sheet (appendix 3), and consent form (appendix 4). At least 24 hours was left before research date to allow time for consideration.

Research Process

Participants attended an in-person one hour VR simulation session (figure 6)



The Simulation

The VR simulation was developed by the technology company Goggleminds^{Ltd} in collaboration with clinicians. The VR device was a HMD Meta Quest 3 with two handheld controllers (figure 7- photo by Painter, 2024).

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Painter, L. (2024, February 15). Meta Quest 3. *Trusted Reviews*. <https://www.trustedreviews.com/reviews/meta-quest-3>

Participants moved around the virtual space and interacted with the environment through a combination of drop-down menus and use of controllers to interact with objects. The participants talked to the patient, completed examinations, performed tests, and gave treatment. The simulation responded to actions taken. For example, if the patient's oxygen level was low, giving oxygen would improve it. Within the scenario it was not possible cause harm, and an integrated verbal cue would be given if a significantly erroneous prescribing action was attempted. However, the patient's clinical situation would not improve without correct action.

Pre Brief and Practice

A pre brief was included to ensure participant practical and psychological safety, with awareness that they could stop at any time. This is in line with best practice simulation guidance (Somerville et al., 2023). Firstly, participants were shown the simulation space and interaction points on a screen. Then a practice simulation allowed interactive familiarisation and practice of the think aloud (TA) technique. The practice scenario concerned an upper gastrointestinal bleed (Version UG1.100002) and was selected due to difference from the research scenario.

Researched Simulation

The scenario (Version SE1.100009) concerned sepsis secondary to a community acquired pneumonia. Sepsis is a serious condition characterised by life threatening organ dysfunction secondary to infection, and is estimated to cause 48,000 deaths annually in the UK (Sepsistrust, 2025). The scenario was selected due to the importance of the topic alongside clear management steps in national guidance (NICE, 2024). I also had personal experience with the scenario, having utilised it extensively in teaching. It has been academically investigated to consider its instructional design and user experience. This indicated high immersion, motivation and learning scores but also moderately high cognitive load (Jacobs et al., 2023).

The simulation concerns a male patient in their 70s presenting with a cough, confusion and feeling unwell (figure 8). He has a low oxygen level, high respiratory rate, high heart rate, low blood pressure, and a low temperature. A chest Xray shows pneumonia, and bloods tests indicate infection and acute kidney injury. The collection of findings should enable the diagnosis of sepsis secondary to a community acquired pneumonia.

The figure originally presented here cannot be made freely available via ORA because of copyright.

The end point of the simulation was determined by one of:

- 1.) Participant decision to handover.
- 2.) Prompt to handover at approximately 18 minutes.
- 3.) Should a participant call for advice early, opportunity was given to complete further actions up to a maximum of 20 minutes.

A short debrief was completed with participants to answer questions, provide feedback and ensure psychological safety in line with simulation standards (ASPiH, 2023).

Data Collection

Aspect One: Actions and Outcomes

An outcome-based method compared participant actions and diagnosis with a model answer. This aspect was selected due to the centrality of outcomes within CR which is ultimately goal orientated to diagnose and treat patients (Daniel et al., 2019). Diagnostic accuracy has also been central to seminal CR research and actions a key outcome measure in previous research into CR in VR simulation.

Model Answer. The model answer consisted of three different areas.

- 1.) **Diagnosis** - based upon verbalised thoughts throughout the scenario or during handover.

The main outcome measure was identification of sepsis, but identification of community acquired pneumonia was also noted. Time of first mention was recorded.

- 2.) **Completion of the Sepsis Six criteria.** These criteria represent six, time critical actions for patients with suspected sepsis, with studies suggesting these reduce mortality (Daniels et al., 2011). This is often taught using the memory aid ‘taking 3 and giving 3’ (GeekyMedics, 2024).

Point of note: The Sepsis Trust created an updated version of the Sepsis Six following new National Institute of Clinical Excellence (NICE) sepsis guidelines (NICE, 2024; Sepsistrust, 2024). However, the new criteria are not fully embedded in education and have broader categories such as ‘monitor’. The earlier version of the Sepsis Six (figure 9, Sepsistrust, 2018; Whittington Health NHS Trust, 2019) was adopted because it offered better clarity in determining actions completed.

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Sepsistrust. (2018). About Sepsis: References & Sources. *The UK Sepsis Trust*. <https://sepsistrust.org/about-sepsis/references-sources/>

3.) A checklist of expected actions was developed from the NICE (2024) sepsis guidelines and adapted for actions possible within VR (appendix 5). Completion of actions was a binary measure with each action weighted equally.

Aspect Two: Sequential Flow

An observational approach was adopted to consider sequential actions taken. This aspect was felt important to consider CR as a longitudinal process rather than just as constituent parts. It was initially considered based on Rochlen et al's (2023) sequential behaviour analysis of CR. However, this is a numerical way to consider behaviour sequences needing high sample sizes. Observation was felt more appropriate within this study.

Video recording of the simulations was performed by casting VR headset input (i.e. from the participant's viewpoint) and recording with Microsoft Teams. A timeline of actions was developed from the video data including menu options selected, practical actions (such as giving oxygen) and review of findings (considered from either explicit selection or from participant gaze).

Aspect Three: Think Aloud

Think Aloud (TA) has been used extensively in CR research. Ericsson & Simon, (1980) suggested TA as a protocol for accessing working memory, as an indirect way to consider more complex internal mental processes. It has been used to research CR alongside classroom activities (R. Patel, 2022; Pinnock et al., 2015; Power et al., 2017) and simulation (Burbach et al., 2015; Elstein et al., 1978). The predominance of this as a research technique within CR highlights its value in considering the unseen reasoning behind actions.

Despite the benefits of TA, it does have limitations. Burbach et al. (2015) used TA alongside standard simulation to investigate CR and found challenges with cognitive overload, and participants narrating actions rather than stating immediate thoughts. Narration is reframed thought and so is not considered true TA. Indeed, some felt TA itself may change the cognitive process, a phenomenon called reactivity (Leow & Morgan-Short, 2004). Although, Fox et al. (2011) performed a meta-analysis showing TA itself does not statistically impact cognition outcomes. However, explaining rather than purely verbalising thoughts was found to improve task performance versus silent controls (Fox et al., 2011). This highlights the need to instruct participants to try to verbalise unmediated thought.

TA was used in this study to attempt to access the black box of the internal mental processes and gain an understanding of participants reasoning and cognitive approach. Recognising the potential pitfalls of this method, tips suggested by Noushad et al. (2024) for using TA to capture cognitive processes were adopted. This included allowing practice, giving prior instruction, minimising interruption and recording TA verbatim.

Audio recording used Microsoft Teams auto transcription. Interruptions were kept to a minimum with the researcher only speaking to encourage to ‘keep speaking your thoughts’ or if there were specific challenges with simulation mechanics. Transcriptions were checked manually afterwards for accuracy and to ensure pauses and non-word utterances (e.g. um) were included.

Analytical Strategy

Baseline Questionnaire

Analysed descriptively regarding the three experience groups to consider participant context.

Integrated Analysis

The three aspects (actions/outcomes, sequential flow, TA) were analysed in an integrated manner. Integrated analysis is essential to mixed methods to make the most of this methodology (Creswell, 2022). Both areas of data convergence and divergence were considered, as the interfaces can lead to greater insights than each source separately (Shannon-Baker, 2016).

Aspect One: Actions and Outcomes

Outcomes of diagnostic accuracy were analysed regarding whether diagnoses were mentioned, time of first mention and whether they were included in final handover. Actions were recorded in a binary manner, with a percentage of actions completed versus adapted NICE (2024) guidelines recorded. Averages per group were utilised, but due to small sample sizes, raw data has been included in results to provide greater transparency.

Time to mention sepsis and percentage of actions completed was compared between experience groupings with a Kruskal Wallis test using the statistical programme R (R version 4.5.1, 2025). Significance was set at $p=0.05$. As no statistically significant differences were identified, a post-hoc test was not required.

Aspect Two: Sequential Flow

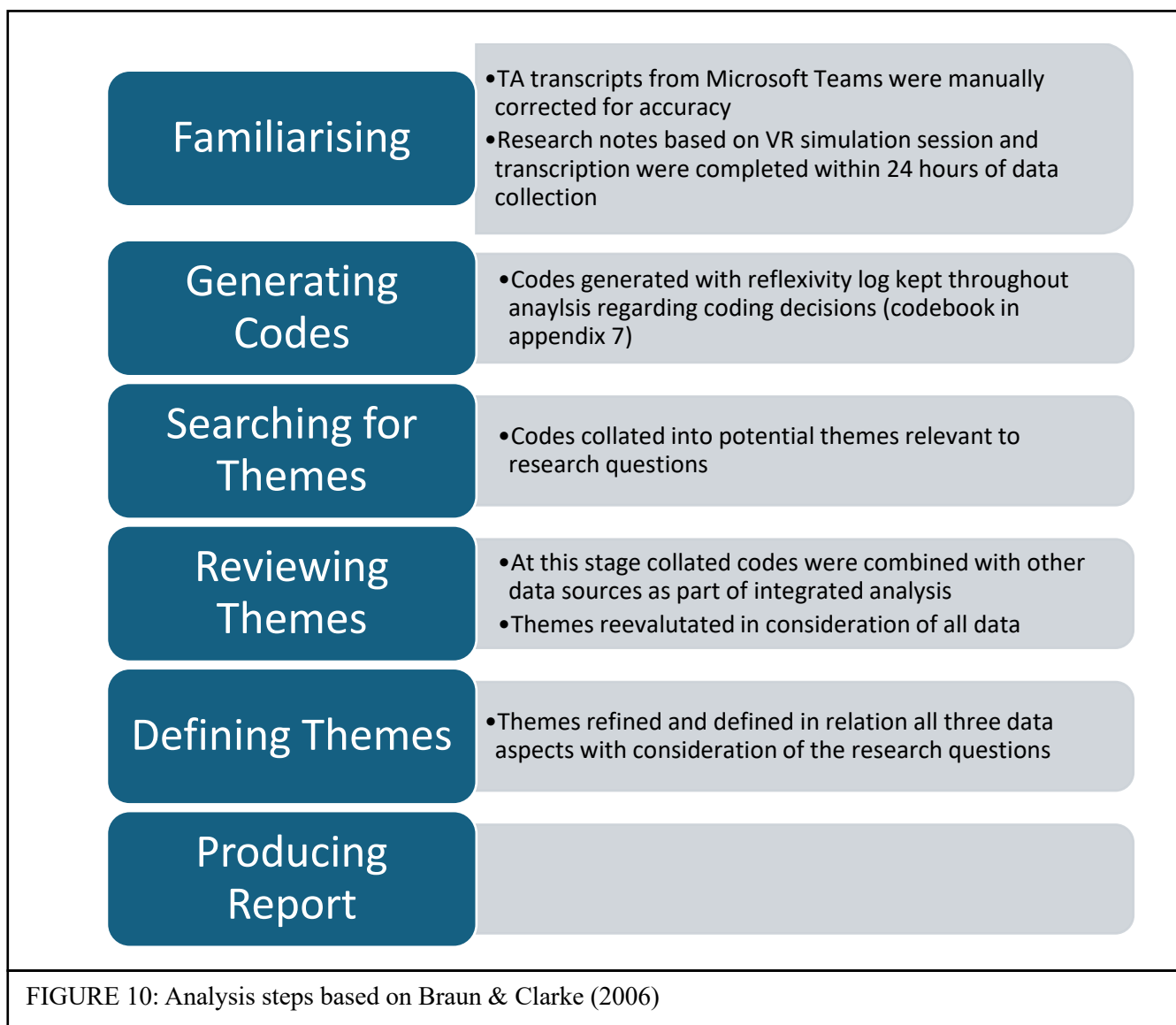
Flow of actions was determined by the researcher in relation to the main activity undertaken by the participant within each 15 seconds of simulation. The main classification of activity is set out in table 2 (further detail in appendix 6). The data was reflected visually for overall description.

Main Type of Activity
Asking Patients Medical History
Examining Patient
Prescribing or Giving Treatment
Ordering/ Reviewing Investigations
Using Medical Guidelines
Handing Over to Another Clinician

TABLE 2: Main activity categories

Aspect Three: Think Aloud

TA was analysed using an adapted form of thematic analysis based on Braun & Clarke's (2006) six phases. To integrate analysis, the other data aspects were included from step four, with themes produced based on the full dataset. Figure 10 sets out the analysis steps.



Quality and Rigour

This methodological approach has both strengths and weaknesses. I considered how to ensure research quality utilising Tracy & Hinrichs' (2017) qualitative criteria and specific rigor advice for mixed methods (Lorenzini et al., 2024). I sought to demonstrate the topic worthiness through my introduction to the literature. I ensured transparency through description of methods and VR context. This aimed to provide credibility and allows readers to consider transferability of findings to their own context. The strength of my approach lies in its triangulation of data sources and methods through integrated analysis. This increases research rigor by providing a more complete picture, cross-verifying evidence and combining objective findings with more subjective observations to reduce bias (Lorenzini et al., 2024).

There were also methodological weaknesses. Firstly, regarding quantitative rigor. Sample sizes were kept small, due to the focus being depth. However, this means that the study lacked statistical power for confirmation of quantitative research trends, reducing reliability (Serdar et al., 2021). The study does not seek generalisability and the sample sizes used are considered appropriate for TA, in which a minimum of five participants is considered adequate (Noushad et al., 2024). Despite this, data saturation was not reached, apart from in the early career analysis group. Data saturation occurs when no new codes are identified in transcripts and is used to assess the adequacy of purposive sample sizes (Hennink & Kaiser, 2022). This would suggest a larger sample size may produce more complete insights. This finding is likely due to each analysis group creating different insights and so a minimum of five per grouping may have been more appropriate.

The other main methodological consideration is the impact of the researcher on the study. This can be particularly challenging in a TA protocol, as researcher interruptions can affect results (Noushad et al., 2024). I acknowledged researcher subjectivity and was considerate of its effect. I used a set research protocol to ensure consistency between participants in my actions. I also kept research notes (appendix 8) within 24 hours of each session and a reflexivity log (appendix 9) throughout analysis. Self-critiquing and evaluating the influence of the researcher helped promote transparency regarding the researcher impact (Olmos-Vega et al., 2023).

Reflexivity Statement

I am a medical doctor and educator by background. I work at the location of my study and have had a role in the early implementation of VR simulation as a teaching resource. I reflect that this may influence a more positive viewpoint regarding this tool. In addition, I had a personal relationship as either colleague or teacher to many of the participants, some senior and some junior to my role. Whilst I tried to reduce the impact of this by using a set research protocol, I found it challenging at times to avoid fulfilling the educator role. A TA protocol required minimal interruption; however, I needed to answer practical questions on how to do tasks within the VR simulation. Whilst answering these, I stuck to practical advice and avoided inadvertently giving scenario assistance. I am also an early career clinician myself, which may have led to my engagement more with those narratives. Additionally, I was conscious of the potential for participants feeling professional jeopardy, especially senior or peer clinicians being observed by a colleague. I ensured to acknowledge these feelings but also clarified the

non-assessment purposes of this study. By keeping research notes and a reflexivity log, I aimed to remain cognisant of my impact as a researcher throughout.

Ethical Considerations

Ethical approval was granted by the Oxford Research Ethics Committee (DREC870605 – appendix 10) as well as local agreement (appendix 11).

Risks. Risks and mitigations (appendix 12) were considered and explained to participants in the information sheet and in person. No adverse events occurred during the study. To protect participant identity, I decided not to report University or speciality as these may make data more identifiable.

Consequences. Potential participants were made aware there was no obligation to take part, that the study would have no bearing on any educational or employment processes and that withdrawal was possible at any point until dissertation submission.

Potential Perceived Conflict of Interest. Goggleminds^{Ltd} provides Great Western Hospital access to their hardware and VR simulation software without charge as part of their product development process. They allowed utilisation for the purposes of this master's project but had no involvement in the data collection, analysis or write up.

Data Storage and Protection. All data was stored on secure university systems under password protection. Participants were given a pseudonymised code used within data from the point of transcription. Data was stored and will be retained or destroyed in line with University of Oxford policy (CUREC, 2020).

Changes and Reflections

The research path is rarely smooth, and some adaptations were made during the process. Three key points were noted:

1. Due to low recruitment numbers, an ethical addendum was sought and approved to expand the recruitment criteria to include clinicians of any experience level leading to three analysis groups.
2. Although, initially, an inductive approach to data coding was considered, overall, this research took an abductive approach to analysis. In abduction, an observation is extrapolated to the most likely, but not only, conclusion often linked to known theory (Veen, 2021). During analysis, it became clear that previous theories may guide some data findings and were appropriate to the coding process, however, new emergent findings were also present. Although an a priori code book was not used, codes linked to theory such as ‘hypothesis testing’ were used alongside inductive codes. Abductive reasoning has been highlighted for its ability to give more realistic account of the messy process of research (Andrade, 2023).

3. Finally, an issue with VR equipment battery life should be noted, which may have influenced one participant's approach (EC1) due to perceived time pressure, though they reported feeling they had adequate time.

Results

Participant Description

Ten participants were recruited. See appendix 13 for full description of recruitment.

Demographics

Gender (self-identified): Four male and six female participants. All analysis groups contained a mix of genders.

Age:

- **Medical student:** All 18-24
- **Early career:** All 25-34
- **Experienced clinician:** Ranged from the 25-34 to >65 category

The baseline demographics showed expected findings with a split of genders in each group and age increasing with expertise. The experienced clinician grouping included consultants and registrars, so the wider age range is expected.

Prior Experience

Group	Years Experience	Standard Simulation Experience (Hours)	VR Simulation Experience (Hours)
Medical Student (MS1)	Early Clinical	0-4	0-4
Medical Student (MS2)	Early Clinical	4-8	0-4
Medical Student (MS3)	Late Clinical	16-24	0-4
Early Career (EC1)	1-2	Over 40	0-4
Early Career (EC2)	3-4	12-16	0-4
Early Career (EC3)	3-4	12-16	0-4
Early Career (EC4)	3-4	16-24	0-4
Experienced (EX1)	5-9	Over 40	0-4
Experienced (EX2)	> 20	0-4	0-4
Experienced (EX3)	> 20	4-8	0-4

TABLE 3: Summary of participant prior experience

Years of experience were reflective of analyses groups, with a wider spread within the experienced clinician group.

Standard simulation experience varied but was higher in a group spanning from MS3 to EX1.

VR experience all participants had minimal exposure to VR simulation.

Overview of Results

To explore the results in relation to both research questions, the mixed methods data will be presented in an integrated manner.

Aim 1: To explore how clinical reasoning can be demonstrated in VR simulation.

Three themes related to the first aim were identified:

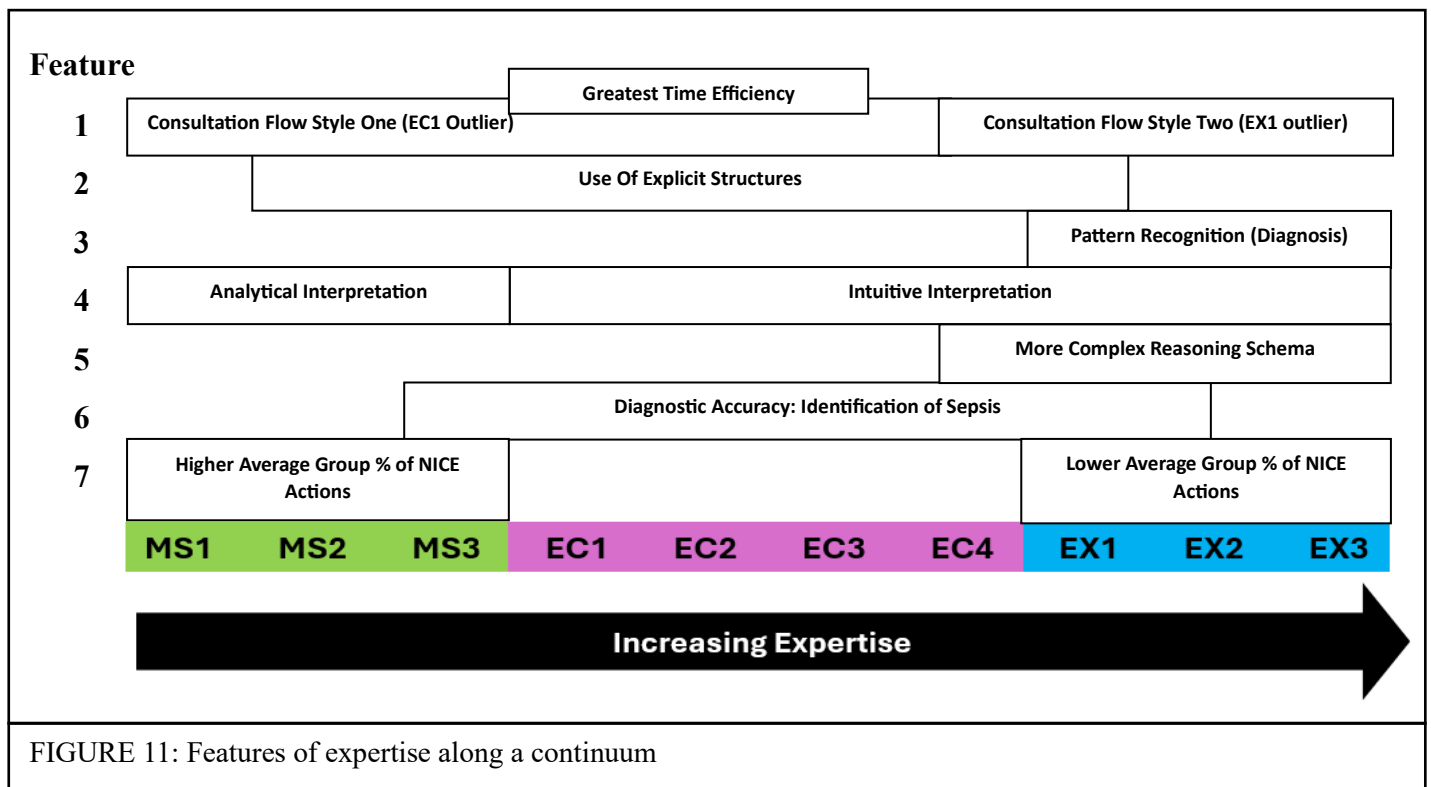
- 1.) Not just theories: how data demonstrated clinical reasoning related to theories and models
- 2.) Indicators of CR errors: common cognitive pitfalls
- 3.) The context: how VR simulation may impact the CR process

Aim 2: To investigate how clinicians of different expertise levels approach clinical reasoning in VR simulation.

The findings did not neatly fit into expertise groupings. However, if expertise is viewed as a continuum, seven features around approach to CR were found. Figure 11 demonstrates them along the spectrum of expertise. These features were not binary with some evidence of each in most transcripts; however, certain features were more prominent depending on expertise. **These features were highly interlinked with the aim one themes so will be integrated throughout to prevent repetition.**

Features along the range of expertise
(these will be indicated through results using a blue box)

- 1.) Differences in overall consultation flow
- 2.) Use of explicit structures
- 3.) Pattern recognition in diagnosis
- 4.) Differences in interpreting investigations
- 5.) Complexity of reasoning schema
- 6.) Diagnostic accuracy – identification of sepsis
- 7.) Differences in average percentage of NICE recommended actions completed



The Overall Flow and Key Actions

The following chart shows an overview, demonstrating the general flow and Sepsis Six actions taken by each participant. It will be referred to within the results.

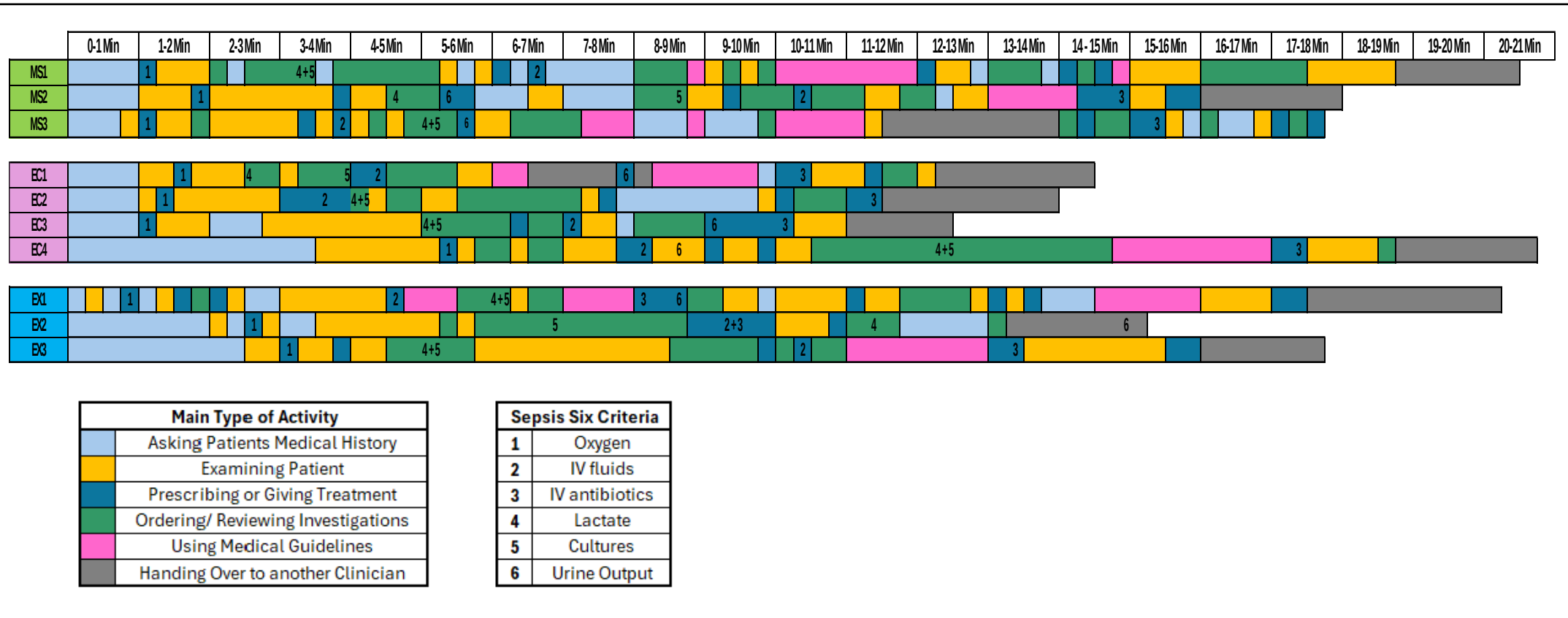


FIGURE 12: Overall simulation flow and key actions by participant

Analysis of Overall Flow and Structure

Expertise Feature One: Differences in overall consultation flow

This data showed the most time efficient group were early career clinicians, generally taking less simulation minutes.

In terms of pattern, all medical students, EC2 and EC3 took a short initial patient history then returned to this later. EC1 took a short patient history and only returned to history later to confirm allergy status. However, EC1 may have felt time pressure due to low battery life, potentially impacting their approach. EC4, EX2 and EX3 spent longer initial time on history and followed a traditional pattern of history, then examination, then investigations. EX1 combined the initial history with taking immediate action but still focused on patient history at the beginning.

Use of A-E Structure

Expertise Feature Two: Use of explicit structures

A to E assessments were developed as a structured way to assess and treat patients (Thim et al., 2012). It focuses on critical issues first, aiming for patient stabilisation rather than seeking a diagnosis. However, A-E was consistently used in this study as a memory aid to structure data gathering and management as part of CR. All apart from MS1, EX2 and EX3 used this structure.

Participants used the A-E structure to situate themselves or as a memory aid to recall key checks:

“So, airways good, breathing, move onto circulation. Do his cap refill” MS2

“Okay, so I’m now starting with my A to E, so A, he’s he’s spoken to me and, um, I didn’t see anything in his mouth. B sats of 90%.” EC1

“Airway’s fine, breathing should reassess just because we’re going through an A to E” EX1

It was also used to rapidly reassesses the patient:

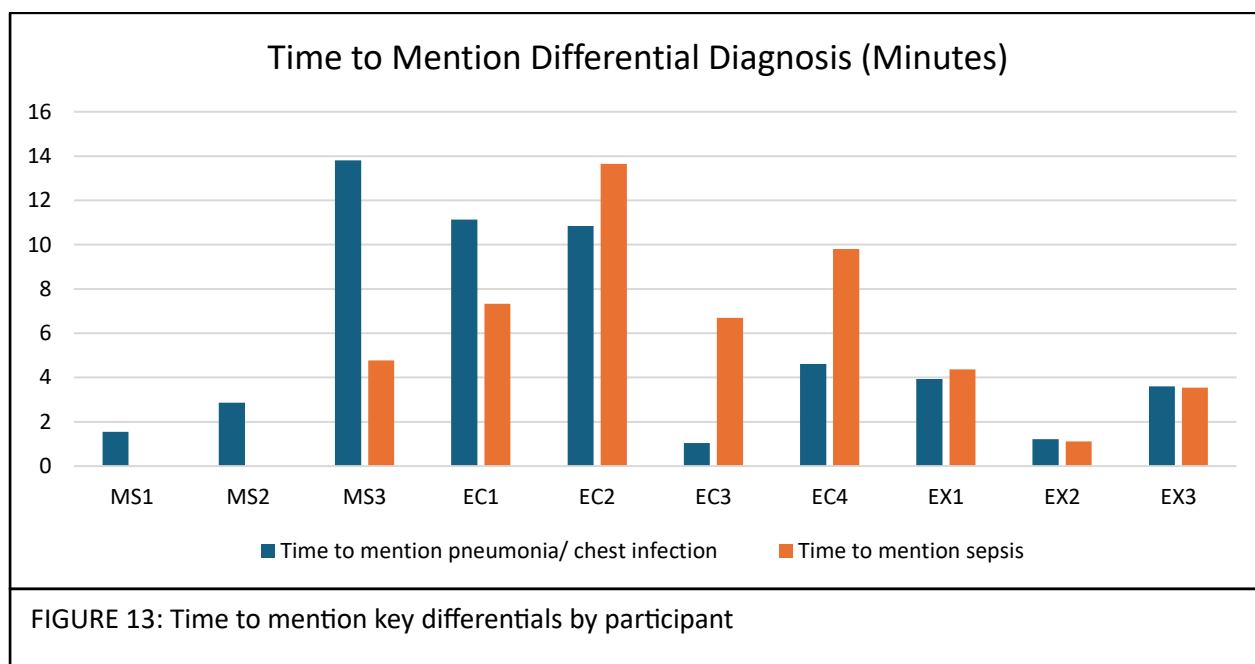
“I’m just quickly going to run through my A to E again before I handover. Airway is patent. He’s talking to me. Breathing, sats have come up. Blood pressure I’m just going to repeat. Um, um he’s got fluids running and I’ve given some antibiotics. Um, D he’s still confused as far as I am aware. Um, I would do pupils. Um and E (looks over patient) Ok, fine” EC4

Theme One: Not Just Theories - How Data Demonstrated Clinical Reasoning Related to Theories and Models

This theme considers how, when viewing data through the lens of previous literature, there was evidence of described theories and models within VR simulation. However, this was not a simple relationship, and multiple reasoning strategies were present in every transcript. There was evidence that CR is possible to demonstrate in VR simulation.

Hypothetico-deductive Reasoning

There was early accurate hypothesis generation. Figure 13 shows **8/10** participants mentioned either pneumonia or sepsis within five minutes.



An initial hypothesis was verbalised after asking the patient questions:

“Ok, so he sounds like he’s got a little bit of a cold, cough, shivering so I’m a bit worried that he’s probably having some sort of chest infection” EC3

“Okay. So OK, so this chap is short of breath and, um, he’s a bit confused so I’m thinking sepsis, I’m wondering if it’s a chest infection” EX3

Or after initial examination:

“He’s hypoxic and, uh, hypotensive and tachycardic and so I’m concerned about sepsis” EC1

“Temperature of 35.2, I’m quite worried this patient is septic” MS3

Two participants (EC1, EC2) did not mention key differentials until handover. Although hypotheses were not explicitly verbalised, their TA and flow data still indicated hypothesis testing.

Hypothesis Testing. Overall flow was non-linear, but all participants initially asked the patient’s medical history, then examined and then arranged investigations, which is suggestive of hypothesis testing.

The hypothesis was verbally linked to or used to justify next steps to test this:

“So, he’s got a cough. I probably want to examine his chest. Uh, listen for infections.” MS1

“So, I did a blood culture. I’d like a lactate. CRP because I’m worried about infection. Lactate is because, again, I’m worried about sepsis.” EC4

In interpreting findings, these were often linked back to the hypothesis:

“OK so. He’s ever so slightly acidotic. With a low bicarbonate. I think this is just further confirming that he is septic” MS3

“Results so chest Xray looks like there(is) opacity in the lower right, the lower right zone. Lower zone of the right lobe. That fits with the overall picture” EC2

Starting from an initial hypothesis, and contextualising investigation choice and interpretation considering this, shows evidence for hypothetico-deductive reasoning.

Pattern Recognition

Expertise Feature Three: Pattern recognition in diagnosis

Experienced clinicians showed evidence of pattern recognition with very rapid identification of potential sepsis from minimal data. The three earliest sepsis mentions were all experienced clinicians: **3 min 56 sec, 1 min 13 sec, 3 min 36 sec.**

Following just two questions at 1 min 7 sec of simulation time EX2 said:

“Right. Ok. So, it sounds like been septic and a bit delirious. Yes, and you’ve got a chest infection.”

However, a Kruskal Wallis test indicated that the difference in time to mention sepsis between groups was not statistically significant. $H(2) = 5.83$ $P = 0.054$

Pattern Recognition in Interpretating Investigations.

Expertise Feature Four: Differences in interpreting investigations

There was difference in interpretation strategy between medical students and more experienced clinicians. Medical students generally described the findings and then concluded using more underlying pathophysiology knowledge.

Medical Students. *“(Reviewing blood gas) He’s slightly acidotic. Um metabolically. Base excess is low. But lactate’s high, so I’m still thinking infection and um ... metabolic acidosis with low base excess so if base excess is low then he’s not got enough ...” MS1*

“Um, but it definitely looks like there’s some sort of consolidation. There (pointing at lower right zone of lung). Ummmm, maybe some up there as well (pointing at right apex of lung on Xray), not too sure, but he definitely seems like he has a lower respiratory tract infection” MS2

Whereas, early career, and experienced clinicians rapidly identified the pathology or normality, and termed it as the conclusion rather than constituent parts.

Early Career. *“Yep, going to check the ECG. All ok.” EC2*

“(reviewing Xray) Right basal consolidation, so that confirms the diagnosis of chest sepsis” EC3

Experienced Clinician. *“(Reviewing ECG) So no ischaemic changes. It looks good in ECG” EX1*

“Okay chest Xray. It shows some, we’ve got a right lower lobe pneumonia.” EX3

There was an exception with EX1 being more descriptive in analysis, however, ECG interpretation would not form part of their normal job role.

“(reviewing ECG) That looks normal. There’s definitely no obvious like ST elevation or anything. I don’t think you can ... that looks like a sinus rhythm. Apart from that I can’t really tell” EX2

Use of Mental Schemas

Conditional Reasoning as a Mental Schema. The main type of reasoning across all transcripts was conditional reasoning, where X finding = Y action.

This was with diagnostic reasoning. “Um, so I’m thinking at the moment, um, infection of some kind, I’d want to check with a chest Xray” MS1

“Fine, so a little bit tachycardic. So maybe we’ll think about doing an ECG” EC4

It was also seen with management reasoning. “That’s 88 (talking about oxygen saturations)” ... “let’s get some oxygen on straight away” MS2

“(participant checks blood pressure) Bit low so I’m going to start some fluid.” EC2

More Complex Conditional Reasoning Schema.

Expertise Feature Five: Complexity of reasoning schema

There was evidence of more complex reasoning schema with increasing experience.

Two experienced clinicians specifically checked the patient’s medical background for COPD, a lung condition which can indicate a lower oxygen target is necessary for safety. This shows reasoning where X findings = Y action UNLESS Z.

“Ok, so short of breath, he’s got sats of 88. So, let’s pop some oxygen on. Before we do that, let’s just make sure he’s not got any COPD” EX1

“So, you’re quite hypoxic there. So, I think we’ll give you some oxygen. Seem a bit confused as well. I’ll just see what your past medical history ...thing with COPD” EX2

There was also consideration of when X could = Y but an additional factor justified not:

“I don’t think that this is, um, a clot. No because I’ve got consolidation on chest Xray so I’m not doing a D-dimer” EC4

“His heart rate is gone down a little bit since he came in, so we won’t give fluids for now” EX3

Or planning next steps if management failed:

“If not adequately resuscitated after 2 litres, at that point we’d give ITU a call” EX1

This more complex CR with increasing experience is well demonstrated in the decision regarding catheterisation. Urine output monitoring can be done without a catheter, but it is less timely and accurate, however catheterisation is invasive and carries risk.

Medical Student. *“(participant places urinary catheter) I think that was clinically relevant because he said he hadn’t gone in quite a while”* MS2

Early Career. *“So, he said he was not passing much urine, so I’d maybe like to do a catheter to measure his output. We could do this sorta, non-invasive. But it’s a shame, but he’s very confused, so he might not tell the nurses when he’s going to the toilet. So maybe we’ll just do a catheter”* EC4

Experienced. *“Fluid balance because he’s not passing urine so. Um, but yeah, keep an eye on this fluid and I don’t know he might need a catheter, I guess he’s not. I do think he’s got quite dehydrated and not drinking very much”* EX2

These quotes demonstrate a development of if/then reasoning schema from MS2 using simple if=then CR, to EC4 weighing up the decision, to EX2 deciding not to catheterise for now but revisit later.

Use of Illness Scripts

Three findings suggested use of illness scripts as part of CR: handover consistency, less script flexibility in junior participant, and unverified assertions in handovers.

Participant handovers, regardless of expertise, showed remarkable consistency in features highlighted regarding enabling conditions, fault and consequences. Table 4 shows features mentioned by at least three participants:

Category	Specific Feature	No of participants who mentioned
Enabling Conditions	Patient demographics	8
	No past medical history	3
Fault	Pneumonia/ chest infection	10
	Sepsis	7
Consequences	Chest Xray findings	10
	Cough	10
	Respiratory exam findings	9
	Physical health observations	8
	Confusion	7
	Shortness of breath	5
	Raised inflammatory markers	4
	Lactate	4
	Urine output	3
	Abnormal renal function	3
TABLE 4: Summary of features mentioned in handover.		

Less Developed Illness Script. An illness script appeared less complex in an early year medical student, leading to initially erroneously excluding a diagnosis:

“Uh he’s got a cough as well, but no temperature, so that’s quite odd. So probably not a pneumonia” MS2

Whereas more experienced medical students and clinicians recognised low temperature as a potential feature of sepsis:

“Temperature of 35.2, I’m quite worried that this patient is septic” MS3

“Ok so he’s a bit cold. That could be a sign of sepsis” EC4

This may represent the more complex illness scripts which develop through experiential learning, creating several potential scripts for disease rather than one textbook description.

The participant who initially discounted pneumonia based on the low temperature was able to adapt their CR with more information, potentially developing their illness scripts.

“White bloods cells are high. Ok, so it, it seems like there is an infection, he’s just not febrile”
MS2

Unverified Assertions. A final piece of evidence for use of illness scripts was an addition of minor unverified assertions to the verbal handover. These were pieces of evidence which would fit with the diagnosis but were not evidenced by true findings in the simulation.

“He is a male over the age of 65 who has a background of a fever” MS2 (no fever)

“Presented with a few days history of shortness of breath and now a productive cough” EX1 (cough never described as productive)

“SpO2 on arrival was about 70 on air” EX3 (SpO2 never below 80%)

“So, he presented with shortness of breath, cough and feeling dizzy” EC2 (no symptom of dizziness)

Although these were minor alterations, participants appeared unaware of their use of nonfactual information. I suggest that all are facts which would be in keeping with illness scripts for pneumonia and/ or sepsis and therefore ended up accidentally linked in handovers.

Dual Process

As demonstrated, both analytical (hypothetico-deductive) and intuitive (pattern recognition) processing were used.

One participant showed dual processing in relation to one investigation. They initially interpreted the chest Xray with an intuitive approach:

“Um, so it looks like he’s got consolidation in the right lung. Um maybe at the bases maybe some kind of pneumonia.” MS1

Then reinterpreted in a more analytical manner:

“So, it does look like there’s some tracheal deviation, but I can’t see a lung collapse and the heart seems to be shifted to this side. Unless this is just the ... increased opacity. Trying to look behind the heart and can’t see very much” MS1

This showed interpretation of normal variants or pneumonia features as additional pathology, despite correct interpretation on initial pattern recognition. Interestingly, within the final handover, they stuck with their more confident initial interpretation of *“consolidation on the lower right lung.”* MS1

Problem Representation

Problem representation was seen both in handover and throughout to summarise key information and refocus, often leading to decision points:

“Um, so I’m thinking that he’s short of breath, he’s got a cough, he’s maybe a little bit feverish, a bit confused. So, I’m thinking about a pneumonia. Um or some other kind of infection.” EC4

“He’s got a high white cell count in keeping with his infection and his chest Xray and his clinical examination both suggest a right lower lobe pneumonia. In keeping with, with everything that we’ve said. Um, so he needs admission, and I will now hand him over to the medical team for, um, for admission for, um chest infection.” EX3

I suggest that successful problem summarising and reframing frees up working memory and allows forward progression.

Theme Two: Indicators of CR Errors - Common Cognitive Pitfalls

This theme considered data which showed evidence of errors in the CR process, where these occurred and any patterns in experience groupings.

Outcomes and Actions

This data gave indications of where CR may not have led to correct diagnosis, investigation, or management.

Diagnosis.

- 10/10 participants identified pneumonia.
- 8/10 participants considered sepsis, however only 6/10 stated sepsis in their final handover.

Expertise Feature Six: Diagnostic accuracy – identification of sepsis

MS1 and MS2 did not mention sepsis throughout. Additionally, EX2 and EX3 considered but did not state sepsis in the final handover. EX2 mentioned sepsis on multiple occasions and fully completed the Sepsis Six, however did not explicitly mention it in the final handover for uncertain reasons. EX3 mentioned sepsis on multiple occasions, however, they stated in their handover that *“I don’t think currently he had evidence of sepsis, but I was concerned about his confusion at home prior to his admission”* EX3.

Although I did not audio record the post simulation debrief, my research notes document that this participant stated the VR patient looked and acted more well than the observations suggested (appendix 8), which may have contributed to discounting sepsis.

Omissions in Investigative and Management Actions.

Expertise Feature Seven: Differences in percentage actions completed of NICE guidelines.

Table 5 shows percentage of completed actions compared to the model answer. All participants performed >70% of actions but there was individual variability. When taken at a group level, there was an inverse relationship between average percentage actions completed and experience level. A Kruskal Wallis test indicated that the difference in percentage actions between groups was not statistically significant. $H(2) = 0.96$ $P = 0.619$

Expertise Level	Average Percentage of Action Completed per group	Individual Percentage of Actions Completed
Medical Students	85.33%	80
		84
		92
Early Career	83%	80
		80
		84
		88
Experienced	80%	88
		80
		72

TABLE 5: Percentage of model answer actions completed based on NICE guidelines (2024)

Completion of Sepsis Six. 100% of actions were completed by 7/10 participants, including one who had not identified sepsis. The most common omission was measuring urine output (table 6).

Sepsis Six Completion							
	Percentage of Sepsis Six Completed (%)	Oxygen	IV fluids	IV antibiotics	Lactate	Blood cultures	Urine output
Medical Students	66.66						
	100						
	100						
Early Career	100						
	83.33						
	100						
	100						
Experienced	100						
	100						
	83.33						

TABLE 6: Completion of Sepsis Six by participant

Five participants who identified sepsis made explicit reference to the Sepsis Six structure, apart from EC2, EX2 and EX3. The Sepsis Six was traditionally taught using the memory aid ‘taking 3 and giving 3’ (GeekyMedics, 2024). These phrases were often referred to.

*“So, I can take three, so I’ve taken blood cultures and I’m going to monitor urinary output” ...
 “Um taken cultures. Um, I’ve given fluid. I’ve taken the, I need, I’ve taken the lactate, taken cultures. I’ve given fluids and need to give antibiotics. Giving oxygen as well.” MS3*

“Um, I’ve given ... So, sepsis six wise, we’re worried about a chest infection” ... “Um, we’ve given oxygen. Sorry, we’ve taken lactate, we’ve taken cultures. Take. We need a urine output. What do we do to catheterise him?” ... “Done catheter, and then we’ve given oxygen and given fluids” EC3

“Fine, so we’ve taken bloods, got a catheter in, given fluids, oxygen, antibiotics. So, we’ve done our sepsis six” EX1

This schema did have some evidence of effectiveness. All participants who used it completed all of the Sepsis Six, whereas 2/3 who identified sepsis but did not use the structure did not complete all six steps. Interestingly, neither of the consultants explicitly used either A-E or Sepsis Six.

Errors in Cognitive Process

These errors were identified from action and TA data.

Errors in Inferences. 3/10 participants did not identify that the patient was confused. This symptom was not explicitly stated but could be inferred.

“Ok, wandering around and getting lost, that doesn’t sound very good, but could indicate confusion” MS3

“Seem a bit confused as well” EX2

As confusion can be a marker for sepsis, this was a potentially important omission.

Errors in Interpretation. These occurred when reviewing data.

“ECG, Ok, so. This isn’t normal, it’s regular, but there aren’t any. There are P waves. They don’t seem to be there (looking at another lead) so maybe there are no p waves” MS1 – ECG was a normal ECG.

“ABG. A normal PH. Low bicarb. Oh, and low PCO2 as well, but PH is normal, so that is ... respiratory ... acidosis with metabolic compensation” MS2 – ABG shows metabolic acidosis with respiratory compensation.

These were seen across expertise groups, but more often in less experienced participants.

Use of Guidelines and Choice of Management

From the flow data, early career clinicians spent noticeably less time on guidelines, with two not reviewing them at all. EX2 also did not view the guidelines. Although this did not prevent safe management of sepsis it did lead to these participants not giving recommended antibiotics.

Scoring Systems as a Decision Aid. CURB65 is a scoring system which indicates severity of pneumonia (Lim et al., 2003). Initially developed as a predictor of mortality and indicator for need of intensive care, it is often used to determine which antibiotics to give.

C - Confusion

U – Urea > 7mmol/L

R – Respiratory rate ≥ 30

B – Blood pressure under 90/60

65 – age over 65

No participants used CURB65 before the guidelines mentioned it, prompting its calculation:

“Of, of course, CURB 65. How could I have forgot that? Right. Algorithms” MS2

This, however, may be authentic, as this type of guideline is used in genuine clinical environments.

Table 7 summarises participants scoring and antibiotic decisions:

Scoring and Antibiotic Decisions				
	CURB65 Score	Identified Confusion	Used Guidelines?	Antibiotics
Medical Student	1			Oral Amoxicillin
	4			IV Amoxicillin
	3			IV co-amoxiclav + oral clarithromycin
Early Career Clinician	3			IV co-amoxiclav
				IV Amoxicillin
				IV Amoxicillin
	3 or 4			IV co-amoxiclav
Experienced Clinician	1			IV Amoxicillin
				IV Amoxicillin
				IV co-amoxiclav

TABLE 7: Calculation of scoring systems and antibiotic decisions by participant

For those who used CURB 65, only two correctly calculated a score of 4, despite all necessary information being present in the simulation. The error cause was inconsistent with some not identifying confusion, some misinterpreting the urea and some not noting the low diastolic blood pressure.

Not using guidelines or miscalculating the CURB65 score was associated with not choosing the guideline advised antibiotic. **6/10** participants did not give advised antibiotics. However, they selected amoxicillin which is the antibiotic of choice for pneumonia in the hospital the study took place within.

The inability to use the scoring systems may be linked to cognitive overload. To calculate the CURB65, participants had to look at several screens in an unfamiliar environment, possibly making it hard to hold all the facts within working memory.

“Uh, he’s not confused, I think. And the urea? Oh wait, urea is in the bloods testing. Look for the urea. That was 28 (highlighting respiratory rate). Done a blood pressure, less than 90? Not anymore. So, he’d be a 1 or 2, I’ll check the urea more than 7. He’d be a 2. Oh, so he’s a CURB one.” MS1

“I’ve lost track of sort of, what I’ve ordered and where it’s going to, what the um, where the results are” EC1

“Let’s just look. So, check the urea. The urea is 10. Confused and urea. Respiratory rate. So, I need to get an updated resp rate. Okay is it greater than 10 (looking at urea) Yes. So, it is ...uh. Age you were born in 47. Is he over 65. Yes, correct? 1968, 1948. Ok so we’ve got a CURB score of three.” MS3

Theme Three: The Context - How VR Simulation may Impact the CR Process

This theme considers how contextual factors within VR simulation may have impacted on CR. This theme focuses on areas where VR may negatively impact CR or increase the difficulty of educators interpreting participants actions.

Technology Specific Issues

EC1 experienced a significant technology issue due to the battery running out in the middle of their final handover.

An additional glitch involved a button to give the patient a blanket not consistently functioning. This led to me needing to intervene with several participants to say “*We will assume that has happened*” which will have reduced immersion.

Fidelity

There were some areas where imperfect fidelity within VR may have impacted CR.

Communication Fidelity. Currently, communication within this VR simulation involves a pre-set list of questions.

One participant liked this scaffolding:

“(looking at questions list) It’s nice that it’s all kind of ... ordered” MS2

However, it limited participants from being able to get the information they felt was important:

“(scrolling questions menu) I don’t think any of these. Let’s see if you want to tell me about anything about confusion” MS3

“And then ... if I ask, about, any urination (flicking through menu of questions) Ummmm, I don’t see anything about urine” MS2

This level of scaffolding questions also led to some participants just asking all the available questions. For example, some asked, ‘Do you have any pain anywhere ‘to which the patient answers no. Followed immediately by ‘What would you rate your pain 0-10’.

Participants commonly scrolled through click-down menus within the VR simulation, seeming to use this as reminders for actions. Sequential actions noted things such as *“checks through clinical tasks menu”* or *“checks through menu of potential questions”* often followed by ordering investigations. In some situations, this is authentic scaffolding, e.g. having a list of blood tests to order from. Nevertheless, this could be considered as ‘over-scaffolding’ and meant participants did not engage with an active CR process.

Examination Fidelity. Although the examinations possible were quite authentic, there are several examinations not currently possible within VR:

“Don’t know if we can check his pupils somehow? Researcher - There isn’t a way to do those at the moment” EC3

“And feel pulse (tries to feel patient pulse) Researcher - So you can’t actually feel a pulse.” EX1

There were periods in all transcripts where I had to intervene to explain how to do a task within VR.

“Um, now I need to listen with the stethoscope. Researcher: If you just keep hold of it with you, yeah. You just need to keep pressing down the middle button while it's. Yeah, there you go.” EX3

“I want to examine him. Uh, where is that, I can't remember, is it this one? (hovering over clinical tasks)” EX2

Investigation Fidelity. Tests were sometimes conducted due to their availability more than indication.

“Right, he's not known to be diabetic, but I feel like I need to do this (talking about blood sugar). Cause it's there. Let's do it quickly” EX2

In places the investigations did not fit with clinical examination, causing participants to question their own CR.

“So, it looks like there's consolidation at the bottom here (pointing at right lower zone of lung) which wouldn't really match with the crackles up here (pointing at patients right upper zone), but I guess that's a code thing” MS2

“So, it's sinus. About 1,2,3,4 (counting squares) 75 beats per minute. Regular. No big ST changes anywhere. He's not as tachy on this” EC3

There was also uncertainty about which investigations may come back within the simulation:

“D-dimer pending. Is that? No. Maybe I didn't request that? No, I thought I did? Researcher - Pending will mean it won't come back within the limits of the scenario” EX2

“(participant requests urine cultures) Is that going to do anything? Researcher - That suggests it won't come back within the limits of the sim” EX1

However, it was impossible to know if expecting results back or to change was evidence of less experiential knowledge, or uncertainty with what to expect in simulation.

One participant stated *“I might repeat the bloods. Now he’s had some fluid.”* And then after checking repeat result *“OH. So, they don’t change when you repeat them.”* MS1

A repeat test this early is not required as you would not expect such quick change. However, the VR simulation is sped up compared to reality. This makes it impossible to tell if something is a reasoning gap or simulation process uncertainty without exploring in a debrief.

Discussion

This mixed methods study performed an in-depth analysis of the CR process of three different expertise groups within a VR simulation. The opportunity to triangulate the sequential flow, outcomes of reasoning, and spoken aloud cognitive processes provided a comprehensive view of CR in VR. This study aimed to investigate how CR can be demonstrated, as well as any variations in approach between different levels of expertise. My discussion summarises the main findings and explores them in depth with relation to the research questions and previous literature.

Aim 1: To explore how clinical reasoning can be demonstrated in VR simulation.

The study results show that CR can be demonstrated in numerous ways within VR simulation. By looking at the data deductively there were features linked to hypothetico-deductive reasoning, knowledge structure theories, and dual process theories. CR was not a straightforward process, and one model alone was insufficient to explain the findings. Although all participants demonstrated sound CR, errors were also identified. These errors were within objective outcomes, such as diagnostic accuracy and participant actions, but also looking at the why behind decision making through TA. There were situations identified where correct outcomes came after incorrect reasoning and imperfect outcomes from sound reasoning. Having a more in-depth understanding provides opportunity for understanding the role of VR simulation in teaching CR. Despite this, there were areas identified where VR simulation may have impacted CR or limited the ability to practice this authentically. The data suggests that, whilst VR simulation is a valuable tool for teaching CR within medical education, it is not suited for all purposes. The following section discusses the findings in relation to the three themes identified and the relevant literature.

Theme One: Not Just Theories - How Data Demonstrated Clinical Reasoning Related to Theories and Models

Evidence supportive of several theories and models regarding CR was visible in the data from the VR simulations.

Hypothetico-deductive Reasoning. Hypotheses were generated early with actions selected and interpreted to test these. Within the sequential flow, all participants followed a pattern of collecting history, followed by examination and then investigations. This is in line with generating and testing hypotheses and consistent with the theory of hypothetico-deductive reasoning suggested by Elstein et al. (1978) and Neufeld et al. (1981). This was a predictable finding, as this theory of reasoning was developed from studies with similar methods: using simulation with TA. Two participants did not explicitly state an early hypothesis, however, their sequential flow and TA transcripts still suggested hypothesis testing. The data corroborates early CR studies but is also open to similar criticisms: that this reasoning did not differentiate between experts and novices.

Pattern Recognition. Pattern recognition was also identified, and this differentiated between expertise levels. Experienced clinicians mention sepsis, the key diagnosis, incredibly early in the simulation, though the trend of rapid recognition was not statistically significant. However, this finding concurs with previous evidence that experts produce faster initial hypothesis (Barrows et al., 1982). Pattern recognition was also seen in interpretation of investigations. TA showed early career and experienced clinicians rapidly interpreted investigations and converted them to diagnosis, whereas medical students adopted a more descriptive approach and considered the underlying pathophysiology.

Conditional Reasoning Schema. Conditional reasoning was frequently demonstrated within all transcripts. These rules-based approaches are a form of mental schema. Almost all management reasoning was forward in approach, where IF (finding) = THEN (action)

Example of forward reasoning “*(participant checks blood pressure) Bit low so I’m going to start some fluid.*” EC2

In contrast, diagnostic reasoning often used backward reasoning where the hypothesis is considered first, working in reverse to determine how to prove it.

Example of backward reasoning “*Um, so I’m thinking at the moment, um, infection of some kind, I’d want to check with a chest Xray*” MS1

V. Patel & Groen (1986) noted that in other fields experts commonly use forward reasoning and novices backward, but this was not apparent within this study. All analysis groups used both forward and backward reasoning.

Illness Scripts. There were three features suggesting use of illness scripts. Firstly, evidence in verbal handovers (table 4). There was high consistency in features highlighted as important. These linked to the three parts of illness scripts: enabling conditions, fault and consequences (Barrows & Feltovich, 1987). Secondly, a medical student's illness script appeared less flexible, with pneumonia excluded based on no fever.

“Uh he’s got a cough as well, but no temperature, so that’s quite odd. So probably not a pneumonia” MS2

With greater experience, clinicians appeared to have more complex illness scripts and did not exclude pneumonia based on one feature. From practice, they likely had exposure to various scripts for pneumonia. This is consistent with a study exploring how experts address situations when the illness script is an imperfect fit (Wheeler et al., 2017). It also may suggest a role for the use of VR simulation in developing illness scripts. With more information, MS2 was able to adapt their reasoning and potentially expanded their mental illness script - *“White bloods cells are high. Ok, so it, it seems like there is an infection, he’s just not febrile” MS2*

The third finding suggesting use of illness scripts was inclusion of unverified assertions within handovers, such as reporting a productive cough or a fever without factual basis. I suggest that this is due to these features being consistent with the textbook illness script for pneumonia. Therefore these symptoms were unwittingly mentioned despite not being present. This has been seen before within medicine and in broader psychological research on use of mental schemas (Arkes & Harkness, 1980; Custers, 2018). Custers (2018) suggested that script typical information (such as fever in my study) is more likely to be falsely recalled.

Dual Process. Intuitive reasoning (pattern recognition) and analytical reasoning (hypothetico-deductive) were both seen, providing evidence for dual process theory as described by Wason & Evans (1974). One medical student initially used an intuitive approach to interpretation, followed by a more analytical approach. The intuitive interpretation was more accurate concurring with studies that analytic thinking may increase false positives, especially by less experienced clinicians (Monteiro et al., 2024). The fact that both intuitive and analytical reasoning were demonstrated is promising, as increased diagnostic accuracy has been found if both strategies are used together (Eva et al., 2007; Norman et al., 1999).

In summary, the first theme suggests that the CR identified within VR simulation can be explained by a similar theoretical framework as reasoning in clinical practice. This adds to the evidence that CR can be authentically practiced within VR, which is essential for this tool being used in medical education.

Theme Two: Indicators of CR Errors - Common Cognitive Pitfalls

The value of VR simulation for teaching and assessment, lies not only in its ability to identify reasoning but also errors. Overall, no participant completed 100% of NICE (2024) recommended actions and **3/10** did not complete the Sepsis Six, actions known to have a mortality benefit (Daniels et al., 2011). Whilst actions suggested error, the TA data allowed for a deeper understanding of the underlying causes. Errors were found in making inferences, interpreting investigations, and calculating scores. Potential causes for these errors are cognitive bias, faulty data gathering or failing to synthesise information.

Forty different kinds of cognitive errors in CR have been described, often linked to cognitive bias (Scott, 2009). Cognitive biases are unconscious errors related to the use of mental shortcuts. An example of cognitive bias is availability bias, where the likelihood of something is estimated based on commonly or recently encountered events (Scott, 2009). This is typically linked with diagnostic error but, in this research, was seen in antibiotic selection with overreliance on previous experience. Participants counted on experiential knowledge, did not consult the simulation guidelines, and prescribed amoxicillin. Whilst amoxicillin is recommended for pneumonia at Great Western Hospital, participants should have consulted the guidelines. NICE (2024) highlights the importance of using local antibiotic guidelines, which reflect the most appropriate antimicrobials in a specific geography. Experiential knowledge has value but overreliance on this can be problematic. Confidence in diagnosis is only mildly correlated with accuracy (Friedman et al., 2005). The data suggests that participants may have felt it was unnecessary to consult guidelines, because this was a frequently encountered situation.

Graber et al. (2005) noted that errors in data gathering or synthesis commonly led to fault. My data identified that incorrect inferences or omissions in history taking led to missing the patient's confusion, a key symptom suggesting sepsis. Even with correct data gathering and interpretation, Groves et al. (2003) found that failure to integrate clinical data led to 20% of error. An example of synthesis failure within this study are decisions regarding antibiotics. To use the guidelines, participants had to calculate a CURB65 score and then select the correct antibiotic. Only one participant achieved this. Since scoring systems and guidelines are recommended as ways to help reduce CR errors (Scott, 2009) this finding is worthy of further consideration.

Whilst six participants attempted the CURB65 calculation, only two correctly synthesised the information. Cowen & Kelley (1994) explored problems in using decision aids, and found data collection mistakes were a sizable source of error. They highlighted the challenge of manually collating data from different sources accurately. To calculate a CURB65 in the VR simulation, participants needed to synthesise information from several screens and pay careful attention where values were near the scoring boundary. This, in addition to TA, appeared to have caused cognitive overload. There were multiple pauses in thought and difficulty consolidating information. In addition, the simulation used has been found to have a moderately high cognitive load (Jacobs et al., 2023). Whilst the VR setting and TA almost certainly increased the cognitive load, genuine clinical environments also have high cognitive load (Laxmisan et al., 2007). The fact that synthesis errors were common could be explained by many factors including the VR setting itself, authentic cognitive overload, or issues with the scoring systems. Further research is needed to investigate this.

In summary, this theme demonstrates that VR simulation can be used to identify not just CR process but also where errors occurred. This is essential for teaching as exploring these errors provides learners opportunities to reflect and develop CR (Forneris et al., 2015). Understanding erroneous reasoning was possible through TA data. Outcomes and actions can only identify that an error occurred but not why. Whilst you would not use TA in most simulations, an expert debrief can fill this role of exploring why actions were taken or omitted (McGaghie et al., 2010). Identifying error as well as CR process makes VR simulation more useful as a teaching tool but could be enhanced with an expert debrief.

Theme Three: The Context - How VR Simulation may Impact the CR Process

CR does not happen in isolation and cannot be divorced from its context. Durning et al. (2013) highlights context as important to CR and describes this using situativity theory (Durning & Artino, 2011). However, in this case the context is not the clinical environment but the simulated one. This theme explores the impact of the VR context on CR in relation to technology limitations, fidelity, and scaffolding.

Technology limitations were evident in limited battery life and buttons not functioning. Additionally, participant unfamiliarity meant the researcher had to explain how to perform actions. External explanations can take away from immersion in the simulation. Immersion has value in enhancing learning (Dede, 2009) by increasing engagement with the scenario (Hagiwara et al., 2016). Whilst the study used a practice simulation to try to limit the impact of unfamiliarity, this would be reduced further with frequent VR simulation use. Direct technology limitations, such as functional errors, are more problematic as anything that reduces the feeling of immersion may distract focus from CR.

Fidelity is another key aspect of simulation-based learning. Fidelity is defined as “*the degree of accuracy to which a simulation, whether it is physical, mental, or both, represents a given frame of reality in terms of cues and stimuli, and permissible interactions.*” (Tun et al., 2015, p.164). Lower fidelity was evident in some VR simulation aspects which may have impacted on participants CR. For example, the inability to do all examinations, such as feeling the pulse or checking the pupils. Additionally, VR does not reflect the overall appearance of an unwell patient which can give important CR cues. The main area of imperfect fidelity impacting CR was incongruency between patient examination and investigation findings. For

example, “*So it looks like there’s consolidation at the bottom here (pointing at right lower zone of lung) which wouldn’t really match with the crackles up here (pointing at patients right upper zone), but I guess that’s a code thing*” MS2. This is not a VR specific phenomenon and has been described in traditional simulation research. Dieckmann et al. (2007) described these as fiction cues which emphasise differences between simulation and clinical cases, such as artificially reducing time to receive results. Within my data, it was sometimes impossible to tell if actions were due to participants interpreting something as a fiction cue, or errors in knowledge or CR. For example, one participant ordered a blood test twice within the scenario; you would not expect the bloods to change in real life. However, was this a clinical knowledge gap or uncertainty with the workings of the simulated environment? This may have implications for using VR simulation to teach CR and may only be determined through debrief.

VR simulation varied significantly from standard simulation and genuine practice in the amount of scaffolding. Scaffolding is an educational technique to provide structured support for learners to develop, often by breaking down or guiding a task (Masava et al., 2022). It developed from Vygotsky's (1978) theories of learning where learners can be guided to develop within a zone of proximal development. Some scaffolding present in VR simulation was authentic to the clinical environment, such as guidelines and scoring systems. These can help support and structure learning, and the data showed greater use of these by novices. However, VR simulation also includes inauthentic scaffolding, such as drop-down menus or question lists. Within medical education, all currently commercially available VR simulations have this scaffolding, not just the researched system. Although this reduces fidelity, Carey & Rossler (2024) caution against the pursuit of higher realism at the expense of learning. Cognitive load appeared to be high and one way to reduce this would be to increase support levels, such as scaffolding, even if this reduces fidelity (Leppink & Duvivier, 2016). Within

CR, a randomised control study found using scaffolding improved medical student diagnostic efficiency (L. Braun et al., 2017). I suggest that this scaffolding may have particular benefit for more junior learners.

Scaffolding can have negatives by reducing cognitive engagement with a CR process. This was demonstrated though participants menu scrolling and selecting all questions regardless of relevance. After the patient reported no pain, some participants still selected the question asking them to rate the pain. Additionally, tests were performed due to availability rather than necessity *“Right, he’s not known to be diabetic, but I feel like I need to do this (talking about blood sugar). Cuz it’s there. Let’s do it quickly” EX2*. Plackett et al. (2021) reported similar findings when exploring CR using virtual patients. They found some medical students struggled to engage with asking a medical history from a virtual patient. One of their participants stated *“So these questions I guess aren’t like very helpful but because they’re there I’m going to ask them anyway”* (Plackett et al., 2021, p.5). This reflects my findings and shows the potential negatives of the scaffolding for CR teaching, which can lead to gamification of the scenario.

In summary, the third theme identified areas where the VR simulation may have directly impacted on the CR process through technology limitations, imperfect fidelity, and scaffolding. This concurs with Durning et al's (2013) view that context can be as important as internal mental factors in CR. The identified challenges are likely to reduce with repeated participation in the VR simulation format and with technology development. Whilst scaffolding may be beneficial to reduce cognitive load and have value for junior learners; there was evidence in this study that it reduced cognitive engagement with the CR process.

Implications: Where Might VR Simulation Fit within CR Education?

To consider study implications, the findings will be discussed in relation to the consensus statement on the five domains of CR (Cooper et al., 2021). Figure 14 demonstrates my conclusions on which CR domains VR simulation is best suited to teach:

The figure originally presented here cannot be made freely available via ORA because of copyright.

The figure was sourced at Cooper, N., Bartlett, M., Gay, S., Hammond, A., Lillicrap, M., Matthan, J., & Singh, M. (2021). Consensus statement on the content of clinical reasoning curricula in undergraduate medical education. *Medical Teacher*, 43(2), 152–159. <https://doi.org/10.1080/0142159X.2020.1842343>

Clinical Reasoning Concepts. The data illustrated that VR simulation demonstrates several elements of CR theories and models, as well as CR errors. Albeit these were evident following analysis rather than on a surface level. If CR processes were made explicit, it could enable learners to consider their own or others reasoning. However, as teaching CR strategies alone has no clear evidence for improving CR, (Sherbino et al., 2014), this may not be the best way to use VR simulation.

History and Physical Examination. The ability of VR simulation to teach this aspect of CR was mixed. Although participants were able to select questions and practice examinations, there were elements that lacked realism. The scaffolding present meant that options were sometimes selected due to their availability rather than through conscious consideration. There was also frustration demonstrated where questions could not be asked, or examinations performed. Nevertheless, key examination findings such as chest crackles were identifiable by all participants. This may be an advantage over standard simulation, where examination of manikins can be affected by distracting mechanical sounds (Dieckmann et al., 2007). Overall, I feel that the real clinical environment, or standardised patient actors, are better suited to teaching this area of CR. However, as VR simulation's fidelity improves this may change.

Choosing and Interpreting Clinical Tests. This was a strong area within VR simulation that was authentic to clinical practice. The sequence and choice of actions could be accurately determined and objectively compared to a model answer. Pairing the actions taken with the reasoning behind them enabled understanding of where correct decisions may be reached through imperfect CR and vice versa. Cooper et al. (2021) suggested that learners must appreciate the need to consider wider clinical findings when interpreting results. Compared to a classroom session on investigations, this simulation enabled contextual understanding and interpretation. A way to expand on the reasons behind errors would increase learning value.

Problem Identification and Management. This is another area that VR simulation seems well suited to. Participants demonstrated problem identification of key diagnosis throughout the simulations. They also employed problem representation, a short summary of the key case features as an abstract representation. Doing this appropriately has been linked to expertise, and is associated with being able to recall relevant illness scripts and increased diagnostic accuracy (Cutrer et al., 2013). Cooper & Frain (2022) advocate for opportunities to effectively practice this skill, however, classroom-based sessions often use case vignettes which already represent the problem. Nendaz et al. (2000) showed a dramatic drop in diagnostic accuracy when cases are approached from an initial symptom versus a summarised vignette. My data showed problem representation in handovers but also at points of transition or uncertainty as a way of reframing. This may have been used by individuals to reduce their own cognitive load, as chunking information can reduce demands on working memory (Thalmann et al., 2019). VR simulation provides opportunities to practice problem representation and identification.

VR simulation enabled the participants to select and review treatment plans as part of management reasoning. Management CR is arguably a more complex cognitive process than diagnostic reasoning, because there is rarely one right answer and it involves contextual consideration, shared decision making and reassessment (D. Cook et al., 2018). The data demonstrated simple management reasoning, for instance starting oxygen when oxygen levels were low. More complex management reasoning was seen which considered multiple or patient specific factors. A pertinent example was the decision on how to monitor urine output, where the risks and benefits of invasive monitoring were considered. Whilst management reasoning can be practiced and demonstrated within VR, feedback on this would be challenging to automate, as there is generally no one correct answer.

Shared Decision Making. VR simulation did not appear well suited to teaching shared decision making. This final domain of CR is integral to effective management reasoning, as decisions are shaped by several contributors, including the patient (D. Cook et al., 2023; Cooper et al., 2021). Within VR simulation, although multi-player function can enable interprofessional working, there are no opportunities to explain and negotiate with the patient. Without significant investment in developing this feature in VR, integrating shared decision making with patients is hard to envisage.

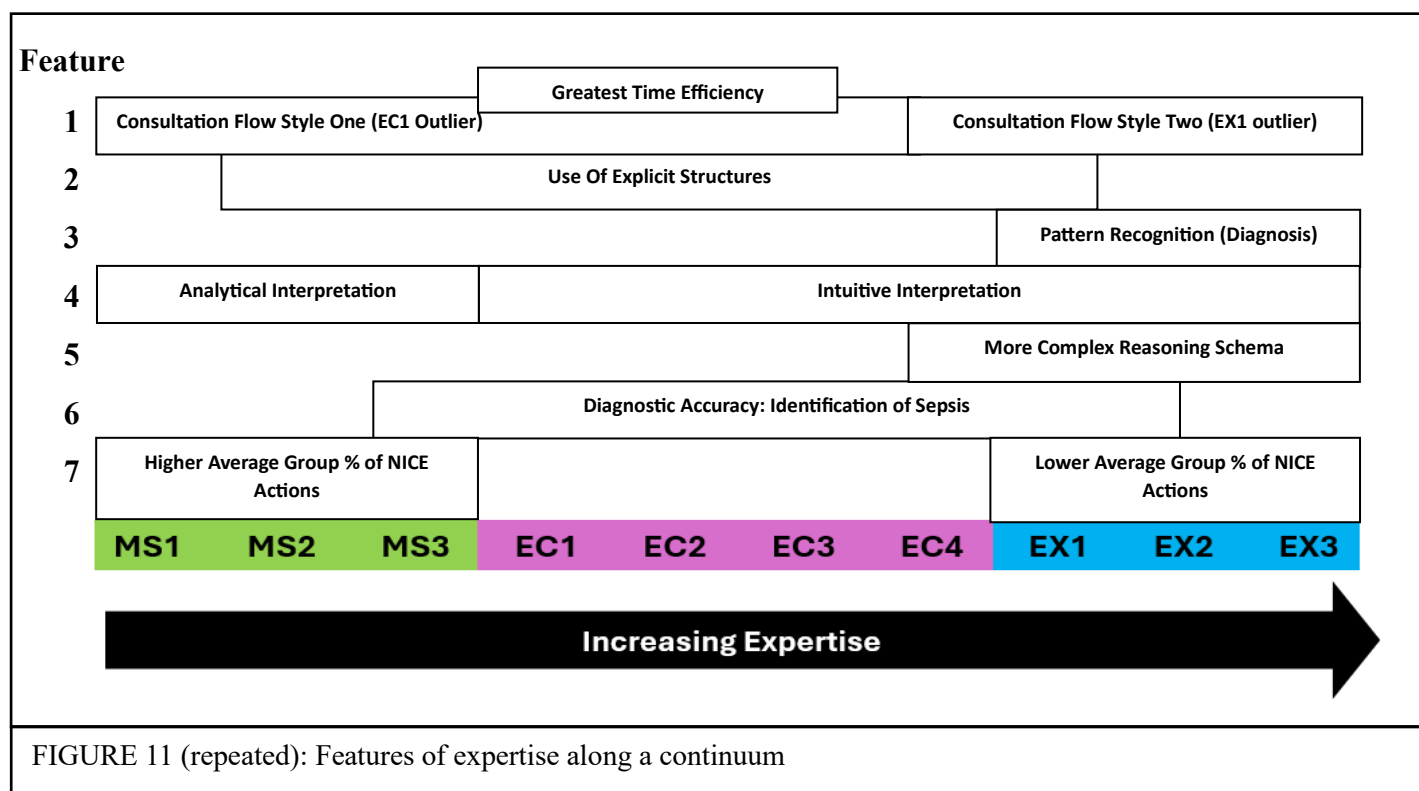
Concluding Aim One

In conclusion, CR can be demonstrated in several ways within VR simulation, and this can be linked to theoretical frameworks. It was possible to investigate both sound CR and errors in a comprehensive way. The TA enabled the outcomes of CR to be linked with the internal mental processes. Drawing on the consensus report (Cooper et al., 2021), the data suggested a role for VR simulation in teaching, but it is better suited to some CR domains than others. These generally reflected the areas of CR which are more science than art. Evidence of VR simulation having a role in CR education may increase learning opportunities. VR enables repeated, adaptable practice (Mergen et al., 2024) and can reduce some barriers associated with traditional simulation (Almoussa et al., 2021). Despite this, the greatest learning may be through a debrief, which is best practice in standard simulation. The data showed that objective outcomes gave a limited view, and deeper understanding required TA. As TA is not used in simulation, a debrief can help make reasoning explicit, so learners can reflect and develop. VR simulation may be an excellent vehicle for learning, but an expert educator debriefing the simulation would amplify the educational benefit.

Aim 2: To investigate how clinicians of different expertise levels approach clinical reasoning in VR simulation.

This study did not seek to generalise findings about differing approaches across expertise levels, but to explore trends regarding this. Seven features relating to expertise were identified and mapped along a continuum.

- 1.) Differences in overall consultation flow
- 2.) Use of explicit structures
- 3.) Pattern recognition in diagnosis
- 4.) Differences in interpreting investigations
- 5.) Complexity of reasoning schema
- 6.) Diagnostic accuracy – identification of sepsis
- 7.) Differences in average percentage of NICE recommended actions completed



The study findings showed divergent conclusions. The three data sources identified different expertise levels as having the most ‘successful’ CR. The findings can be considered in relation to the three data aspects used, to further consider the divergence. These are sequential flow, TA, and outcomes/actions.

Sequential Flow

Feature One: Differences in Overall Consultation Flow. An analysis of the sequential flow demonstrated differences in time efficiency, as well as overall patterns of consultation, between participants. Early career clinicians were typically the most time efficient, with three of this group completing the simulation in the shortest time. Although EC1 and EX1 differed slightly, two general patterns were seen. Most junior participants took a short initial history and then proceeded to other tasks before returning to history later. Whereas more experienced clinicians spent longer on history followed by examination and then investigation. This fits with traditional consultation models and the often quoted study that nearly 80% of diagnosis is found from the patient’s history (Peterson et al., 1992).

Despite different patterns, no one pathway was more like to lead to a correct diagnosis. This contrasts with Rochlen et al's (2023) analysis of sequential behaviour in extended reality which found clear associations between action sequences and successful outcomes. However, that study investigated a larger group with a small number of clear decision points. It may be possible to consider association between action sequence and successful sepsis management within a larger study. Nevertheless, managing a patient case rarely has one correct path, so there is uncertainty regarding whether this would lead to a set of clear actionable strategies for educators.

Feature Two: Use of Explicit Structures. Intermediately experienced participants all used explicit mental structures such as A-E and Sepsis Six to organise their approach and as an aide-memoire. Interestingly, the two most experienced participants did not use either of these approaches. This could be attributed to several factors. Firstly, with increasing expertise, these structures may become superfluous, such as how experienced drivers change gear without conscious thought. Alternatively, the more experienced clinicians may not have been taught these strategies. Whilst A-E assessments were developed in the 1970s, it has been more widely recommended in the UK since 2007 ('ABCDE Approach', n.d.; Thim et al., 2012). In this study, use of A-E correlated with the increased exposure to simulation that participants ranging from MS3 to EX1 had. Use of these structures did appear to be beneficial, with all participants who used the Sepsis Six completing all expected actions. However, it is challenging to draw a strong conclusion from this, as one participant who did not identify sepsis also completed all the Sepsis Six.

The Think Aloud

The qualitative transcripts allowed deeper understanding of the reasoning behind decisions and actions. This concurred with Pardue et al's (2023) purely qualitative exploration of CR in VR which used retrospective interviews to investigate reasoning behind clinical decisions. Considering expertise, pattern recognition was frequently demonstrated by experts regarding diagnosis and investigation. Experts also demonstrated more complex reasoning schemas and nuanced decision making.

Feature Three: Pattern Recognition in Diagnosis. Sepsis was most rapidly identified as a hypothesis by the three most senior clinicians suggesting the use of pattern recognition. This concurs with literature suggesting pattern recognition is more commonly used by experts (Case et al., 1988).

Feature Four: Differences in Interpreting Investigations. Early career and experienced clinicians generally used pattern recognition in interpreting investigations. In contrast, medical students were more descriptive of investigations with greater focus on underlying pathophysiology. Consistent with wider literature, this finding suggests the underlying biological process is given less consideration with increasing expertise (Custers et al., 1996, 1998).

Feature Five: Complexity of Reasoning Schema. More complex reasoning schemas were seen with increasing expertise. For example, clinicians considered past medical history before starting oxygen. One experienced clinician also considered steps beyond initial management such as intensive care. Junior participants schemas occasionally appeared less developed. Sometimes single symptoms were used to discount a diagnosis (e.g. no fever = no pneumonia) whereas more experienced participants did not use the information in isolation. It has been suggested that experts have a larger repository of illness scripts from experiential learning (Custers, 2015). Therefore, it follows that with increasing expertise, clinicians would have numerous pneumonia examples to draw on and an illness script that does not always require a fever for that diagnosis.

Outcomes and Actions

Feature Six: Diagnostic Accuracy – Identification of Sepsis. Whilst all participants identified pneumonia, only **8/10** mentioned sepsis. The two most junior participants did not consider sepsis within their initial hypothesis or throughout. However, whilst the most senior two participants considered it, they did not include it in the final handover.

Feature Seven: Differences in Average Percentage of NICE Recommended Actions Completed. The average percentage of actions completed compared to NICE (2024) sepsis guidelines showed an inverse relationship with expertise level. There was significant individual variation, and this trend was not statistically significant. However, this finding concurred with the diagnostic accuracy, which suggested increasing expertise did not correlate with greater success in objective measures of CR. Keicher et al. (2024) similarly found that only 65.6% of expected actions were performed correctly when investigating VR emergency scenarios. They concluded this was due to participants being junior, however they did not compare to experts, so it is challenging to draw conclusions.

It would have been reasonable to predict greater diagnostic accuracy and action completion with increasing expertise. I suggest three potential reasons for the data not demonstrating this: the intermediate effect, scenario selection and the VR setting itself.

Firstly, the intermediate effect, a phenomenon where those with intermediate experience outperform experts in simulation-based assessments (Schmidt et al., 1988; Schuwirth, 2009). This was demonstrated by early career participants, who all concluded sepsis

as a diagnosis in their final handover, in contrast to only 1/3 for both medical student and experienced clinicians. Interestingly, the medical students completed the highest average percentage of correct actions. Whilst not statistically significant, this trend runs counter to percentage actions equalling diagnostic accuracy. This could suggest an issue with using checklists as a measure. Norman (2005a) suggests that checklists typically reward thoroughness rather than expertise. Previous studies found that seniors perform worse than juniors on checklist metrics (Hodges et al., 1999). It has been argued that this is due to checklists not taking into account efficiency of data usage that comes with experience (Cox, 1990). However, the checklist used in the study was developed from evidence based NICE guidelines, these factors are important in clinical care and so omitting them is not meaningless. Despite this, the data suggests that a checklist is not the most comprehensive assessment tool for CR. If every button was clicked, the score could be 100%, without engaging in any CR.

The data could also be explained by the VR scenario itself. Outside of emergency medicine and anaesthetics, initial management of acutely unwell patients is rarely the role of an expert, especially out of hours, where early career clinicians are the first responders (Church et al., 2023). This may explain why early career clinicians performed best at identifying sepsis. It should be noted that none of the experienced clinicians within this study worked in anaesthetics or emergency medicine. In contrast, Colonna et al. (2022) found clear correlation between expertise and completion of correct actions in a VR trauma simulation. Their experts, however, were trauma surgeons, therefore, the scenario was role authentic. The sepsis scenario selection could explain why the intermediate clinicians appeared to outperform the experts, but there is no way to directly evidence this within my study.

The final explanation for experts not scoring as highly on outcomes and action metrics relates to the context of VR itself. My research notes record a debrief discussion with one experienced clinician. They felt that the VR patient did not appear to be or act unwell in comparison to the objective findings (appendix 8). Potentially, an end of the bed pattern recognition of a sick patient forms part of expert CR but was not represented as authentically in VR. This would contrast with Zackoff et al's (2020) study, which found better recognition of end of bed signs like respiratory distress in VR simulation. Although, their comparator group used simulation manikins rather than patients. Regardless, clinical gestalt, or more intuitive gut feelings are known to be powerful. A meta-analysis showed that gut intuition was equivalent or better than objective score for 16 different conditions (Dale et al., 2019). Regarding sepsis, Knack et al. (2024) found that clinician gestalt outperformed objective measures for early identification. Gestalt is associated with expertise and is thought to develop through experiential learning, however, it is also considered more at risk of cognitive bias errors (C. Cook, 2009). Despite this, it is clearly used in practice. J. Patel et al. (2024) proposed the question that if it outperforms standardised metrics, then how can curriculums be designed to develop gestalt. Within this study, it is possible that the VR setting did impact the findings as the imperfect fidelity means that clinical gestalt could not be fully employed. If sick patients could be more authentically represented in VR this has the potential for allowing opportunities to develop gestalt, 'gut feeling', and be advantageous for teaching CR over manikin simulations.

Overall, the reasoning for experts scoring less well on objective measures is complex, and this study can make no definitive conclusion. The impact of the scenario selection and imperfect fidelity may explain the findings. However, the most likely explanation is the intermediate effect, which has been well documented in wider literature.

Implications for Assessment of CR

Considering my findings, I suggest caution if VR simulation were to be used in assessment of CR. The repeatability and consistency may improve reliability. Nevertheless, the metric used to determine successful CR risks not truly reflecting this complex phenomenon. If objective outcome measures are taken in isolation without consideration of the why behind actions, VR risks rewarding thoroughness rather than sound CR.

Limitations

Whilst this research conducted a thorough examination of CR, its limitations stem from its small-scale, which restricted the ability to discriminate between expertise levels. It also only investigated one stimulation scenario, which is problematic when CR is known to vary between cases. Given the intention to perform an in-depth analysis of CR rather than broad generalisability, the research could have benefited from the involvement of an additional researcher regarding data analysis to enhance robustness. Additionally, although there is significant crossover between different VR simulations, this study's findings are specific to Goggleminds^{Ltd}'s VR programme. Further studies regarding differences in expertise groups and of numerous VR programmes would be beneficial.

Another limitation relates to convenience sampling with use of medical education channels to advertise. This meant two things: firstly, I had a personal relationship with most participants which may have impacted the research process. I attempted to mitigate this by maintaining a consistent research protocol and by keeping a reflexivity log during analysis. Secondly, the participants were all associated with medical education, this may limit the transferability of findings. Individuals with interest and experience in teaching CR and simulation may have different approaches to a wider population of learners.

Conclusion

This mixed methods study facilitated an in-depth exploration of CR in VR simulation, how it can be demonstrated and differences in approach between clinicians of varying expertise. CR can be demonstrated within VR simulation in numerous ways, underpinned by theoretical frameworks. To the researcher's knowledge, this is the first study to take a detailed view combining process and outcomes of CR in VR simulation. This comprehensive review adds to the growing evidence base for VR simulation having a role within CR education and where it may be best suited to teach.

VR simulation demonstrated a strong potential for teaching CR in two domains: choosing and interpreting investigations, problem identification and management (Cooper et al., 2021). However, the context of the VR simulation itself impacted on the CR process, particularly due to over-scaffolding. Although this may reduce cognitive load, and benefit more junior learners, it also leads to less engagement with the reasoning process. This, together with limited communication options, makes it a less useful teaching tool for history, and shared decision making. Due to its increased scaffolding, VR simulation has particular benefits for novices, who need to first learn the rules before figuring out when these should be broken. Overall, VR simulation is better placed for teaching the science rather than the art of CR.

Triangulating the sequential flow, outcomes and TA data enabled a look into the black box behind clinical decisions. This revealed both sound CR but also potential erroneous reasoning. The qualitative TA data enabled identification of key areas of error which suggests the ongoing need to pair VR simulation with a debrief. Focusing on the why behind actions

would allow full exploration of CR and make the most of learning opportunities. VR simulation is a great vehicle for CR learning but would be optimised using an expert debrief.

It was not possible to precisely differentiate between experience groupings. The data regarding differences in approach between expertise levels was divergent and limited by small study numbers. Different metrics highlighted 'better' performance for the three groups. Medical students performed, on average, the highest percentage of best practice actions. Intermediately experienced clinicians had the best overall outcomes in terms of diagnostic accuracy and efficiency of time. Experienced clinicians demonstrated more complex CR and considered wider factors within decision making. There are several potential reasons for the divergence of the data, which warrant further study. Caution is needed if VR simulation is utilised in assessment as the use of different metrics can result in contrasting findings.

Ultimately, this research confirms that there is a role for VR simulation in CR education. Erroneous CR causes significant mortality and morbidity worldwide, underscoring the need for a new educational tool. VR simulation is not a golden bullet for CR, but it can help our future healthcare professionals to practice this vital skill. In the resource and time constrained educational environment, VR offers significant benefit in allowing the practice of CR. Educationalists need to take a pivotal role in the development of future VR alongside commercial entities, to ensure pedagogical and theoretical considerations remain central. VR simulation does not replace clinical exposure, it compliments it. This tool teaches the science of CR, enabling learners to make the most of clinical environments to learn the art.

References

- Aalten, C. M., Samson, M. M., & Jansen, P. a. F. (2006). Diagnostic errors; the need to have autopsies. *The Netherlands Journal of Medicine*, 64(6), 186–190.
- ABCDE Approach. (n.d.). *Elearning for Healthcare*. Retrieved 1 September 2025, from <https://www.e-lfh.org.uk/programmes/abcde-approach/>
- Abulfaraj, M. M., Jeffers, J. M., Tackett, S., & Chang, T. (2021). Virtual Reality vs. High-Fidelity Mannequin-Based Simulation: A Pilot Randomized Trial Evaluating Learner Performance. *Cureus*, 13(8), e17091. <https://doi.org/10.7759/cureus.17091>
- Al-Elq, A. H. (2010). Simulation-based medical teaching and learning. *Journal of Family and Community Medicine*, 17(1), 35–40. <https://doi.org/10.4103/1319-1683.68787>
- Almoussa, O., Zhang, R., Dimma, M., Yao, J., Allen, A., Chen, L., Heidari, P., & Qayumi, K. (2021). Virtual Reality Technology and Remote Digital Application for Tele-Simulation and Global Medical Education: An Innovative Hybrid System for Clinical Training. *Simulation & Gaming*, 52(5), 614–634. <https://doi.org/10.1177/10468781211008258>
- Al-Saud, L. M., Mushtaq, F., Allsop, M. J., Culmer, P. C., Mirghani, I., Yates, E., Keeling, A., Mon-Williams, M. A., & Manogue, M. (2017). Feedback and motor skill acquisition using a haptic dental simulator. *European Journal of Dental Education*, 21(4), 240–247. <https://doi.org/10.1111/eje.12214>
- Andrade, A. D. (2023). *Chapter 16: Dancing between theory and data: abductive reasoning*. <https://www.elgaronline.com/edcollchap/book/9781802205398/book-part-9781802205398-23.xml>
- Arkes, H. R., & Harkness, A. R. (1980). Effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 568–575. <https://doi.org/10.1037/0278-7393.6.5.568>
- ASPiH. (2023). *ASPiH Standards 2023 – ASPiH*. <https://aspih.org.uk/standards-2/>
- Baidoo, K., & Adu, C. (2025). Virtual reality training to enhance clinical competence and student engagement in Ghana. *British Journal of Midwifery*, 33(1), 6–12. <https://doi.org/10.12968/bjom.2024.0037>
- Barrows, H., & Felton, P. (1987). The clinical reasoning process. *Medical Education*, 21(2), 86–91. <https://doi.org/10.1111/j.1365-2923.1987.tb00671.x>
- Barrows, H., Norman, G., Neufeld, V., & Feightner, J. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine. Medecine Clinique Et Experimentale*, 5(1), 49–55.
- Barteit, S., Lanfermann, L., Bärnighausen, T., Neuhaus, F., & Beiersmann, C. (2021). Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. *JMIR Serious Games*, 9(3), e29080. <https://doi.org/10.2196/29080>
- Bartlett, S. F. C. (1995). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.
- Bonnin, C., Pejoan, D., Ranvial, E., Marchat, M., Andrieux, N., Fourcade, L., & Perrochon, A. (2023). Immersive virtual patient simulation compared with traditional education for clinical reasoning: A pilot randomised controlled study. *Journal of Visual Communication in Medicine*, 46(2). <https://doi.org/10.1080/17453054.2023.2216243>

- Braun, L., Zottmann, J., Adolf, C., Lottspeich, C., Then, C., Wirth, S., Fischer, M., & Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical Education*, 51(11), 1118–1126. <https://doi.org/10.1111/medu.13355>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brentnall, J., Thackray, D., & Judd, B. (2022). Evaluating the Clinical Reasoning of Student Health Professionals in Placement and Simulation Settings: A Systematic Review. *International Journal of Environmental Research and Public Health*, 19(2), 936. <https://doi.org/10.3390/ijerph19020936>
- Burbach, B., Barnason, S., & Thompson, S. A. (2015). Using “Think Aloud” to Capture Clinical Reasoning during Patient Simulation. *International Journal of Nursing Education Scholarship*, 12(1), 1–7. <https://doi.org/10.1515/ijnes-2014-0044>
- Carey, J. M., & Rossler, K. (2024). The How When Why of High Fidelity Simulation. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK559313/>
- Case, S., Swanson, D., & Stillman, P. (1988). Evaluating diagnostic pattern recognition: The psychometric characteristics of a new item format. *Research in Medical Education : Proceedings of the ... Annual Conference. Conference on Research in Medical Education*, 27. <https://pubmed.ncbi.nlm.nih.gov/3218869/>
- Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & van der Vleuten, C. (2000). The Script Concordance Test: A Tool to Assess the Reflective Clinician. *Teaching and Learning in Medicine*, 12(4), 189–195. https://doi.org/10.1207/S15328015TLM1204_5
- Charlin, B., Tardif, J.-C., & Boshuizen, H. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine : Journal of the Association of American Medical Colleges*, 75(2), 182–190. <https://doi.org/10.1097/00001888-200002000-00020>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Church, H. R., Murdoch-Eaton, D., & Sandars, J. (2023). Under- and post-graduate training to manage the acutely unwell patient: A scoping review. *BMC Medical Education*, 23, 146. <https://doi.org/10.1186/s12909-023-04119-1>
- Colonna, A., Robbins, R., Stefanucci, J., Creem-Regeh, S., Patterson, P., Engel, B., Fischer, A., & Nirula, R. (2022). Trauma bay virtual reality-A game changer for ATLS instruction and assessment. *The Journal of Trauma and Acute Care Surgery*, 93(3). <https://doi.org/10.1097/TA.0000000000003569>
- Combs, C. D., & Combs, P. F. (2019). Emerging Roles of Virtual Patients in the Age of AI. *AMA Journal of Ethics*, 21(2). <https://doi.org/10.1001/amajethics.2019.153>
- Cook, C. (2009). Is Clinical Gestalt Good Enough? *The Journal of Manual & Manipulative Therapy*, 17(1), 6–7. <https://doi.org/10.1179/106698109790818223>
- Cook, D., Sherbino, J., & Durning, S. (2018). Management Reasoning: Beyond the Diagnosis. *JAMA*, 319(22), 2267–2268. <https://doi.org/10.1001/jama.2018.4385>
- Cook, D., Stephenson, C. R., Gruppen, L., & Durning, S. (2023). Management Reasoning: Empirical Determination of Key Features and a Conceptual Model. *Academic Medicine: Journal of the Association of American Medical Colleges*, 98(1), 80–87. <https://doi.org/10.1097/ACM.0000000000004810>

- Cook, D., & Triola, M. (2009). Virtual patients: A critical literature review and proposed next steps. *Medical Education*, 43(4). <https://doi.org/10.1111/j.1365-2923.2008.03286.x>
- Cooper, N., Bartlett, M., Gay, S., Hammond, A., Lillicrap, M., Matthan, J., & Singh, M. (2021). Consensus statement on the content of clinical reasoning curricula in undergraduate medical education. *Medical Teacher*, 43(2), 152–159. <https://doi.org/10.1080/0142159X.2020.1842343>
- Cooper, N., & Frain, J. (2022). *ABC of Clinical Reasoning*. John Wiley & Sons.
- Cowen, J. S., & Kelley, M. A. (1994). Errors and Bias in Using Predictive Scoring Systems. *Critical Care Clinics*, 10(1), 53–72. [https://doi.org/10.1016/S0749-0704\(18\)30144-1](https://doi.org/10.1016/S0749-0704(18)30144-1)
- Cox, K. (1990). No Oscar for OSCA. *Medical Education*, 24(6), 540–545. <https://doi.org/10.1111/j.1365-2923.1990.tb02672.x>
- Creswell, J. W. (2022). *A concise introduction to mixed methods research* (Second edition.). SAGE Publishing.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third edition). SAGE.
- Croskerry, P. (2009). Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. *Advances in Health Sciences Education*, 14(1), 27–35. <https://doi.org/10.1007/s10459-009-9182-2>
- CUREC. (2020). Title: *Management and Protection of Data Collected for Research Purposes—Best Practice Guidance 09_Version 4.7*. Central University Research Ethics Committee (University of Oxford). [chrome-extension://efaidnbmnnnibpcajpcgclefindmkaj/https://researchsupport.admin.ox.ac.uk/files/bpg09datacollectionandmanagementpdf](https://researchsupport.admin.ox.ac.uk/files/bpg09datacollectionandmanagementpdf)
- Curran, V. R., Xu, X., Aydin, M. Y., & Meruvia-Pastor, O. (2022). Use of Extended Reality in Medical Education: An Integrative Review. *Medical Science Educator*, 33(1), 275–286. <https://doi.org/10.1007/s40670-022-01698-4>
- Custers, E. (2015). Thirty years of illness scripts: Theoretical origins and practical applications. *Medical Teacher*, 37(5), 457–462. <https://doi.org/10.3109/0142159X.2014.956052>
- Custers, E. (2018). Training Clinical Reasoning: Historical and Theoretical Background. In O. ten Cate, E. Custers, & S. Durning (Eds.), *Principles and Practice of Case-based Clinical Reasoning Education: A Method for Preclinical Students*. Springer. <http://www.ncbi.nlm.nih.gov/books/NBK543756/>
- Custers, E., Boshuizen, H., & Schmidt, H. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & Cognition*, 24(3), 384–399. <https://doi.org/10.3758/bf03213301>
- Custers, E., Boshuizen, H., & Schmidt, H. (1998). The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results From an Interview Study. *Cognition and Instruction*, 16(4), 367–398. https://doi.org/10.1207/s1532690xc1604_1
- Cutrer, W., Sullivan, W. M., & Fleming, A. E. (2013). Educational Strategies for Improving Clinical Reasoning. *Current Problems in Pediatric and Adolescent Health Care*, 43(9), 248–257. <https://doi.org/10.1016/j.cppeds.2013.07.005>
- Dale, A. P., Marchello, C., & Ebell, M. H. (2019). Clinical gestalt to diagnose pneumonia, sinusitis, and pharyngitis: A meta-analysis. *British Journal of General Practice*, 69(684), e444–e453. <https://doi.org/10.3399/bjgp19X704297>

- Daniel, M., Rencic, J., Durning, S., Holmboe, E., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R. J., Sergio Da Silva, A., Cleary, T., Stojan, J., & Gruppen, L. (2019). Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Academic Medicine*, 94(6), 902. <https://doi.org/10.1097/ACM.0000000000002618>
- Daniels, R., Nutbeam, T., McNamara, G., & Galvin, C. (2011). The sepsis six and the severe sepsis resuscitation bundle: A prospective observational cohort study. *Emergency Medicine Journal: EMJ*, 28(6), 507–512. <https://doi.org/10.1136/emj.2010.095067>
- Dede, C. (2009). Immersive Interfaces for Engagement and Learning. *Science*, 323(5910), 66–69. <https://doi.org/10.1126/science.1167311>
- Delavari, S., Monajemi, A., Baradaran, H. R., Myint, P. K., Yaghmaei, M., & Soltani Arabshahi, S. K. (2020). How to develop clinical reasoning in medical students and interns based on illness script theory: An experimental study. *Medical Journal of the Islamic Republic of Iran*, 34, 9. <https://doi.org/10.34171/mjiri.34.9>
- Dieckmann, P., Manser, T., Wehner, T., & Rall, M. (2007). Reality and fiction cues in medical patient simulation: An interview study with anesthesiologists. *Journal of Cognitive Engineering and Decision Making*, 1(2), 148–168. <https://doi.org/10.1518/155534307X232820>
- Donathan, E., LaLumia, A., Foat, C., Barr, N., & Page, D. (2023). *SMARTSIM* A multicenter prospective randomized trial of 3D virtual reality versus traditional patient simulation. *Journal of the American College of Emergency Physicians Open*, 5(1). <https://doi.org/10.1002/emp2.13092>
- Durning, S., & Artino, A. (2011). Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. *Medical Teacher*, 33(3), 188–199. <https://doi.org/10.3109/0142159X.2011.550965>
- Durning, S., Artino, A., Schuwirth, L., & van der Vleuten, C. (2013). Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. *Academic Medicine: Journal of the Association of American Medical Colleges*, 88(4), 442–448. <https://doi.org/10.1097/ACM.0b013e3182851b5b>
- Elstein, A. S. (1994). What goes around comes around: Return of the hypothetico-deductive strategy. *Teaching and Learning in Medicine*, 6(2), 121–123. <https://doi.org/10.1080/10401339409539658>
- Elstein, A. S. (2009). Thinking about diagnostic thinking: A 30-year perspective. *Advances in Health Sciences Education: Theory and Practice*, 14 Suppl 1, 7–18. <https://doi.org/10.1007/s10459-009-9184-0>
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical Problem Solving: A Ten-Year Retrospective. *Evaluation & the Health Professions*, 13(1), 5–36. <https://doi.org/10.1177/016327879001300102>
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251.
- Eva, K. (2005). What every teacher needs to know about clinical reasoning. *Medical Education*, 39(1), 98–106. <https://doi.org/10.1111/j.1365-2929.2004.01972.x>
- Eva, K., Hatala, R., Leblanc, V., & Brooks, L. (2007). Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Medical Education*, 41(12), 1152–1158. <https://doi.org/10.1111/j.1365-2923.2007.02923.x>

- Evans, J., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Feltovich, P., & Barrows, H. (1984). Issues of generality in medical problem solving. In H. Schmidt & M. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*. (pp. 128–142). Van Gorcum.
- Ferrandini Price, M., Escribano Tortosa, D., Nieto Fernandez-Pacheco, A., Perez Alonso, N., Cerón Madrigal, J. J., Melendreras-Ruiz, R., García-Collado, Á. J., Pardo Rios, M., & Juguera Rodriguez, L. (2018). Comparative study of a simulated incident with multiple victims and immersive virtual reality. *Nurse Education Today*, 71, 48–53. <https://doi.org/10.1016/j.nedt.2018.09.006>
- Forneris, S. G., Neal, D. O., Tiffany, J., Kuehn, M. B., Meyer, H. M., Blazovich, L. M., Holland, A. E., & Smerillo, M. (2015). Enhancing Clinical Reasoning Through Simulation Debriefing: A Multisite Study. *Nursing Education Perspectives*, 36(5), 304. <https://doi.org/10.5480/15-1672>
- Foronda, C. L. (2021). What Is Virtual Simulation? *Clinical Simulation In Nursing*, 52, 8. <https://doi.org/10.1016/j.ecns.2020.12.004>
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137(2), 316–344. <https://doi.org/10.1037/a0021663>
- Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G. C., Wolf, F. M., Heckerling, P. S., Fine, P. L., Miller, T. M., & Elstein, A. S. (2005). Do Physicians Know When Their Diagnoses Are Correct? Implications for Decision Support and Error Reduction. *Journal of General Internal Medicine*, 20(4), 334. <https://doi.org/10.1111/j.1525-1497.2005.30145.x>
- GeekyMedics. (2024). *Sepsis | Acute Management | ABCDE | Geeky Medics*. <https://geekymedics.com/acute-management-of-sepsis/>
- Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135–146. <https://doi.org/10.1057/ejis.2011.54>
- Gordon, D., Rencic, J., Lang, V., Thomas, A., Young, M., & Durning, S. (2022). Advancing the assessment of clinical reasoning across the health professions: Definitional and methodologic recommendations. *Perspectives on Medical Education*, 11(2). <https://doi.org/10.1007/S40037-022-00701-3>
- Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, 165(13), 1493–1499. <https://doi.org/10.1001/archinte.165.13.1493>
- Groves, M., O'Rourke, P., & Alexander, H. (2003). The clinical reasoning characteristics of diagnostic experts. *Medical Teacher*, 25(3), 308–313. <https://doi.org/10.1080/0142159031000100427>
- Gruppen, L. D. (2016). Clinical Reasoning: Defining It, Teaching It, Assessing It, Studying It. *Western Journal of Emergency Medicine*, 18(1), 4. <https://doi.org/10.5811/westjem.2016.11.33191>
- Gutiérrez, F., Pierce, J., Vergara, V., Coulter, R., Saland, L., Caudell, T., Goldsmith, T., & Alverson, D. (2007). The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. *Studies in Health Technology and Informatics*, 125. <https://pubmed.ncbi.nlm.nih.gov/17377256/>

- Hagiwara, M. A., Backlund, P., Söderholm, H. M., Lundberg, L., Lebram, M., & Engström, H. (2016). Measuring participants' immersion in healthcare simulation: The development of an instrument. *Advances in Simulation*, 1, 17. <https://doi.org/10.1186/s41077-016-0018-x>
- Hartstein, A. J., Zimney, K., Verkuyl, M., Yockey, J., & Berg-Poppe, P. (2022). Virtual Reality Instructional Design in Orthopedic Physical Therapy Education: A Randomized Controlled Trial. *Journal of Physical Therapy Education*. <https://doi.org/10.1097/JTE.0000000000000221>
- Havola, S., Haavisto, E., Mäkinen, H., Engblom, J., & Koivisto, J.-M. (2021). The Effects of Computer-Based Simulation Game and Virtual Reality Simulation in Nursing Students' Self-evaluated Clinical Reasoning Skills. *CIN: Computers, Informatics, Nursing*, 39(11), 725. <https://doi.org/10.1097/CIN.0000000000000748>
- Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
- Higgs, J. (1992). Developing Clinical Reasoning Competencies. *Physiotherapy*, 78(8), 575–581. [https://doi.org/10.1016/S0031-9406\(10\)61202-3](https://doi.org/10.1016/S0031-9406(10)61202-3)
- Hodges, B., Regehr, G., McNaughton, N., Tiberius, R., & Hanson, M. (1999). OSCE checklists do not capture increasing levels of expertise. *Academic Medicine: Journal of the Association of American Medical Colleges*, 74(10), 1129–1134. <https://doi.org/10.1097/00001888-199910000-00017>
- Huhn, K., Gilliland, S., Black, L., Wainwright, S., & Christensen, N. (2019). Clinical Reasoning in Physical Therapy: A Concept Analysis. *Physical Therapy*, 99(4). <https://doi.org/10.1093/ptj/pzy148>
- Ilgen, J., Bowen, J., McIntyre, L., Banh, K., Barnes, D., Coates, W., Druck, J., Fix, M., Rimple, D., Yarris, & Eva, K. (2013). Comparing diagnostic performance and the utility of clinical vignette-based assessment under testing conditions designed to encourage either automatic or analytic thought. *Academic Medicine : Journal of the Association of American Medical Colleges*., 88(10), 1545–1551. <https://doi.org/10.1097/ACM.0b013e3182a31c1e>
- Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, 27(1), 10–28. <https://doi.org/10.1080/01421590500046924>
- Jacobs, C., Vaidya, K., Old, T., & Medwell, L. (2023). A9 Immersive Technology Experience Measure (ITEM): Pilot study on participant experience using novel questionnaire and VR scenario. *International Journal of Healthcare Simulation*, 3(1), Article 1. <https://doi.org/10.54531/RWXN3896>
- Jay, R., Davenport, C., & Patel, R. (2024). Clinical reasoning—The essentials for teaching medical students, trainees and non-medical healthcare professionals. *British Journal of Hospital Medicine*, 85(7), 1–8. <https://doi.org/10.12968/hmed.2024.0052>
- Jayasundera, M., Myers, M., Pandian, K., & Gingell, G. (2022). Virtual Reality Simulation: Evaluating an Experiential Tool for the Clinical Application of Pathophysiology. *Medical Science Educator*, 32(6). <https://doi.org/10.1007/s40670-022-01683-x>
- Jeong, Y., Lee, H., & Han, J.-W. (2022). Development and evaluation of virtual reality simulation education based on coronavirus disease 2019 scenario for nursing students: A pilot study. *Nursing Open*, 9(2), 1066–1076. <https://doi.org/10.1002/nop2.1145>
- Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a Research Paradigm and Its Implications for Social Work Research. *Social Sciences*, 8(9), 255. <https://doi.org/10.3390/socsci8090255>

- Keicher, F., Backhaus, J., König, S., & Mühling, T. (2024). Virtual reality for assessing emergency medical competencies in junior doctors – a pilot study. *International Journal of Emergency Medicine*, 17(1), 125. <https://doi.org/10.1186/s12245-024-00721-2>
- Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 4(6), 533–550. <https://doi.org/10.1111/j.1745-6924.2009.01164.x>
- Kim, S., Lee, Y., & Go, Y. (2023). Constructing a Mixed Simulation With 360° Virtual Reality and a High-Fidelity Simulator: Usability and Feasibility Assessment. *Computers, Informatics, Nursing : CIN*, 41(8). <https://doi.org/10.1097/CIN.0000000000000976>
- Knack, S. K. S., Scott, N., Driver, B. E., Prekker, M. E., Black, L. P., Hopson, C., Maruggi, E., Kaus, O., Tordsen, W., & Puskarich, M. A. (2024). Early Physician Gestalt Versus Usual Screening Tools for the Prediction of Sepsis in Critically Ill Emergency Patients. *Annals of Emergency Medicine*, 84(3), 246–258. <https://doi.org/10.1016/j.annemergmed.2024.02.009>
- Laxmisan, A., Hakimzada, F., Sayan, O. R., Green, R. A., Zhang, J., & Patel, V. L. (2007). The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *International Journal of Medical Informatics*, 76(11–12), 801–811. <https://doi.org/10.1016/j.ijmedinf.2006.09.019>
- Lee, E., Baek, G., & Hwang, Y. (2023). Effectiveness of the Patient's Severity Classification Competency Promotion Virtual Reality Program of Nursing Students during the COVID-19 Pandemic Period. *Healthcare (Basel, Switzerland)*, 11(8). <https://doi.org/10.3390/healthcare11081122>
- Leow, R. P., & Morgan-Short, K. (2004). To Think Aloud or Not to Think Aloud: The Issue of Reactivity in SLA Research Methodology. *Studies in Second Language Acquisition*, 26(1), 35–57. <https://doi.org/10.1017/S0272263104026129>
- Leppink, J., & Duvivier, R. (2016). Twelve tips for medical curriculum design from a cognitive load theory perspective. *Medical Teacher*, 38(7), 669–674. <https://doi.org/10.3109/0142159X.2015.1132829>
- Lim, W. S., van der Eerden, M. M., Laing, R., Boersma, W. G., Karalus, N., Town, G. I., Lewis, S. A., & Macfarlane, J. T. (2003). Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. *Thorax*, 58(5), 377–382. <https://doi.org/10.1136/thorax.58.5.377>
- Lorenzini, E., Osorio-Galeano, S. P., Schmidt, C. R., & Cañon-Montañez, W. (2024). Practical Guide to Achieve Rigor and Data Integration in Mixed Methods Research. *Investigacion y Educacion En Enfermeria*, 42(3), e02. <https://doi.org/10.17533/udea.iee.v42n3e02>
- Macauley, K., Brudvig, T. J., Kadakia, M., & Bonneville, M. (2017). Systematic Review of Assessments That Evaluate Clinical Decision Making, Clinical Reasoning, and Critical Thinking Changes After Simulation Participation. *Journal of Physical Therapy Education*, 31(4), 64. <https://doi.org/10.1097/JTE.000000000000011>
- Mandin, H., Jones, A., Woloschuk, W., & Harasym, P. (1997). Helping students learn to think like experts when solving clinical problems. *Academic Medicine : Journal of the Association of American Medical Colleges*, 72(3), 173–179. <https://doi.org/10.1097/00001888-199703000-00009>
- Masava, B., Nyoni, C. N., & Botma, Y. (2022). Scaffolding in Health Sciences Education Programmes: An Integrative Review. *Medical Science Educator*, 33(1), 255–273. <https://doi.org/10.1007/s40670-022-01691-x>

- McGaghie, W., Issenberg, S., Petrusa, E., & Scalese, R. (2010). A critical review of simulation-based medical education research: 2003-2009. *Medical Education*, 44(1). <https://doi.org/10.1111/j.1365-2923.2009.03547.x>
- McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The quality of health care delivered to adults in the United States. *The New England Journal of Medicine*, 348(26), 2635–2645. <https://doi.org/10.1056/NEJMsa022615>
- Memon, M. A., Thurasamy, R., Ting, H., & Cheah, J. (2025). Purposive Sampling: A Review and Guidelines for Quantitative Research. *Journal of Applied Structural Equation Modeling*, 9(1), 01–23. [https://doi.org/10.47263/JASEM.9\(1\)01](https://doi.org/10.47263/JASEM.9(1)01)
- Mergen, M., Graf, N., & Meyerheim, M. (2024). Reviewing the current state of virtual reality integration in medical education—A scoping review. *BMC Medical Education*, 24(1), 788. <https://doi.org/10.1186/s12909-024-05777-5>
- Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*, E77-D, no. 12, 1321–1329.
- Mills, B., Dykstra, P., Hansen, S., Miles, A., Rankin, R., Hopper, L., Brook, L., & Bartlett, D. (2020). Virtual Reality Triage Training Can Provide Comparable Simulation Efficacy for Paramedicine Students Compared to Live Simulation-Based Scenarios. *Prehospital Emergency Care*, 24(4). <https://doi.org/10.1080/10903127.2019.1676345>
- Monteiro, S., Sherbino, J., LoGiudice, A., Lee, M., Norman, G., & Sibbald, M. (2024). The influence of viewing time on visual diagnostic accuracy: Less is more. *Medical Education*, 58(7), 858–868. <https://doi.org/10.1111/medu.15380>
- Morgan, D. L. (2013). *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach*. SAGE Publications.
- Nendaz, M., & Perrier, A. (2012). Diagnostic errors and flaws in clinical reasoning: Mechanisms and prevention in practice. *Swiss Medical Weekly*, 142(4344), Article 4344. <https://doi.org/10.4414/sm.w.2012.13706>
- Nendaz, M., Raetzo, M., Junod, A., & Vu, N. V. (2000). Teaching Diagnostic Skills: Clinical Vignettes or Chief Complaints? *Advances in Health Sciences Education: Theory and Practice*, 5(1), 3–10. <https://doi.org/10.1023/A:1009887330078>
- Neufeld, V., Norman, G., Feightner, J., & Barrows, H. (1981). Clinical problem-solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education*, 15(5), 315–322. <https://doi.org/10.1111/j.1365-2923.1981.tb02495.x>
- NICE. (2024, 19). *Overview | Suspected sepsis: Recognition, diagnosis and early management | Guidance | NICE*. NICE. <https://www.nice.org.uk/guidance/ng51>
- Norman, G. (2005a). Editorial – Checklists vs. Ratings, the Illusion of Objectivity, the Demise of Skills and the Debasement of Evidence. *Advances in Health Sciences Education*, 10(1), 1–3. <https://doi.org/10.1007/s10459-005-4723-9>
- Norman, G. (2005b). Research in clinical reasoning: Past history and current trends. *Medical Education*, 39(4), 418–427. <https://doi.org/10.1111/j.1365-2929.2005.02127.x>
- Norman, G., Brooks, L., Colle, C., & Hatala, R. (1999). The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning. *Cognition and Instruction*, 17(4), 433–448. https://doi.org/10.1207/S1532690XC11704_3

- Norman, G., & Eva, K. (2010). Diagnostic error and clinical reasoning. *Medical Education*, 44(1), 94–100. <https://doi.org/10.1111/j.1365-2923.2009.03507.x>
- Norman, G., Pelaccia, T., Wyer, P., & Sherbino, J. (2024). Dual process models of clinical reasoning: The central role of knowledge in diagnostic expertise. *Journal of Evaluation in Clinical Practice*, 30(5), 788–796. <https://doi.org/10.1111/jep.13998>
- Norman, G., Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., Kreuger, S., & Monteiro, S. (2014). The etiology of diagnostic errors: A controlled trial of system 1 versus system 2 reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, 89(2), 277–284. <https://doi.org/10.1097/ACM.000000000000105>
- Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: The role of experience. *Medical Education*, 41(12), 1140–1145. <https://doi.org/10.1111/j.1365-2923.2007.02914.x>
- Noushad, B., Van Gerven, Pascal W. M., & de Bruin, A. B. H. (2024). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, 46(7), 892–897. <https://doi.org/10.1080/0142159X.2023.2289847>
- Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2023). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher*, 45(3), 241–251. <https://doi.org/10.1080/0142159X.2022.2057287>
- Painter, L. (2024, February 15). Meta Quest 3. *Trusted Reviews*. <https://www.trustedreviews.com/reviews/meta-quest-3>
- Palmas, F., & Klinker, L. (2020). *Defining Extended Reality Training: A Long-Term Definition for All Industries*. 322–324. <https://doi.org/doi:10.1109/ICALT49669.2020.00103>
- Panda, S. C. (2006). Medicine: Science or Art? *Mens Sana Monographs*, 4(1), 127–138. <https://doi.org/10.4103/0973-1229.27610>
- Pansiri, J. (2005). Pragmatism: A methodological approach to researching strategic alliances in tourism. *Tourism and Hospitality Planning & Development*, 2(3), 191–206. <https://doi.org/10.1080/14790530500399333>
- Pardue, K., Holt, K., Dunbar, D., & Baugh, N. (2023). Exploring the Development of Nursing Clinical Judgment Among Students Using Virtual Reality Simulation. *Nurse Educator*, 48(2). <https://doi.org/10.1097/NNE.0000000000001318>
- Patel, J., Jasti, J., & Gavinski, K. (2024). Demystifying Gestalt. *Annals of Emergency Medicine*, 84(6), 696–697. <https://doi.org/10.1016/j.annemergmed.2024.06.027>
- Patel, R. (2022). General practice trainees' learning experiences of formative think-aloud script concordance testing. *Education for Primary Care*, 33(4), 229–236. <https://doi.org/10.1080/14739879.2022.2057240>
- Patel, V., & Groen, G. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10(1), 91–116. [https://doi.org/10.1016/S0364-0213\(86\)80010-6](https://doi.org/10.1016/S0364-0213(86)80010-6)
- Pelaccia, T., Tardif, J., Tribby, E., & Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: The dual-process theory. *Medical Education Online*, 16, 10.3402/meo.v16i0.5890. <https://doi.org/10.3402/meo.v16i0.5890>
- Penne, J., & Anderson, C. (2020). Adventures in Teaching Clinical Reasoning. *The Journal of Physician Assistant Education*, 31(1), 48. <https://doi.org/10.1097/JPA.0000000000000294>

- Peterson, M. C., Holbrook, J. H., Hales, D. V., Smith, N. L., & Staker, L. V. (1992). Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Western Journal of Medicine*, 156(2), 163.
- Pinnock, R., Young, L., Spence, F., Henning, M., & Hazell, W. (2015). Can Think Aloud Be Used to Teach and Assess Clinical Reasoning in Graduate Medical Education? *Journal of Graduate Medical Education*, 7(3), 334–337. <https://doi.org/10.4300/JGME-D-14-00601.1>
- Plackett, R., Kassianos, A. P., Mylan, S., Kambouri, M., Raine, R., & Sheringham, J. (2022). The effectiveness of using virtual patient educational tools to improve medical students' clinical reasoning skills: A systematic review. *BMC Medical Education*, 22(1), 365. <https://doi.org/10.1186/s12909-022-03410-x>
- Plackett, R., Kassianos, A. P., Timmis, J., Sheringham, J., Schartau, P., & Kambouri, M. (2021). Using Virtual Patients to Explore the Clinical Reasoning Skills of Medical Students: Mixed Methods Study. *Journal of Medical Internet Research*, 23(6), e24723. <https://doi.org/10.2196/24723>
- Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, 6(3), 181–185. <https://doi.org/10.7861/fhj.2019-0036>
- Powell, T. C. (2001). Competitive advantage: Logical and philosophical considerations. *Strategic Management Journal*, 22(9), 875–888. <https://doi.org/10.1002/smj.173>
- Power, A., Lemay, J.-F., & Cooke, S. (2017). Justify Your Answer: The Role of Written Think Aloud in Script Concordance Testing. *Teaching and Learning in Medicine*, 29(1), 59–67. <https://doi.org/10.1080/10401334.2016.1217778>
- R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rencic, J. (2011). Twelve tips for teaching expertise in clinical reasoning. *Medical Teacher*, 33(11), 887–892. <https://doi.org/10.3109/0142159X.2011.558142>
- Rochlen, L. R., Putnam, E. M., Tait, A. R., Du, H., & Popov, V. (2023). Sequential Behavioral Analysis: A Novel Approach to Help Understand Clinical Decision-Making Patterns in Extended Reality Simulated Scenarios. *Simulation in Healthcare*, 18(5), 321. <https://doi.org/10.1097/SIH.0000000000000686>
- Schmidt, H., Boshuizen, H., & Hobus, P. (1988). Transitory Stages in the Development of Medical Expertise: The “Intermediate Effect” in Clinical Case Representation Studies 1. In *10th Annual Conference Cognitive Science Society Pod*. Psychology Press.
- Schmidt, H., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Medical Education*, 49(10), 961–973. <https://doi.org/10.1111/medu.12775>
- Schmidt, H., Norman, G., & Boshuizen, H. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine: Journal of the Association of American Medical Colleges*, 65(10), 611–621. <https://doi.org/10.1097/00001888-199010000-00001>
- Schuwirth, L. (2009). Is assessment of clinical reasoning still the Holy Grail? *Medical Education*, 43(4), 298–300. <https://doi.org/10.1111/j.1365-2923.2009.03290.x>
- Scott, I. A. (2009). Errors in clinical reasoning: Causes and remedial strategies. *BMJ*, 338, b1860. <https://doi.org/10.1136/bmj.b1860>
- Senvisky, J. M., McKenna, R. T., & Okuda, Y. (2024). Financing And Funding A Simulation Center. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK568786/>

- Sepsistrust. (2018). About Sepsis: References & Sources. *The UK Sepsis Trust*. <https://sepsistrust.org/about-sepsis/references-sources/>
- Sepsistrust. (2024). Clinical tools. *The UK Sepsis Trust*. <https://sepsistrust.org/healthcare-professionals/clinical-tools/>
- Sepsistrust. (2025). About sepsis. *The UK Sepsis Trust*. <https://sepsistrust.org/about-sepsis/>
- Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, 31(1), 010502. <https://doi.org/10.11613/BM.2021.010502>
- Shannon-Baker, P. (2016). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research*, 10(4), 319–334. <https://doi.org/10.1177/1558689815575861>
- Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., Kreuger, S., & Norman, G. (2012). The relationship between response time and diagnostic accuracy. *Academic Medicine: Journal of the Association of American Medical Colleges*, 87(6), 785–791. <https://doi.org/10.1097/ACM.0b013e318253acbd>
- Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *CJEM*, 16(1), 34–40. <https://doi.org/10.2310/8000.2013.130860>
- Sim, J., Rusli, K., Seah, B., Levett-Jones, T., Lau, Y., & Liaw, S. (2022). Virtual Simulation to Enhance Clinical Reasoning in Nursing: A Systematic Review and Meta-analysis. *Clinical Simulation in Nursing*, 69. <https://doi.org/10.1016/j.ecns.2022.05.006>
- Simmons, B. (2010). Clinical reasoning: Concept analysis. *Journal of Advanced Nursing*, 66(5), 1151–1158. <https://doi.org/10.1111/j.1365-2648.2010.05262.x>
- Somerville, S., Harrison, N., & Lewis, S. (2023). Twelve tips for the pre-brief to promote psychological safety in simulation-based education. *Medical Teacher*, 45(12), 1349–1356. <https://doi.org/10.1080/0142159X.2023.2214305>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Tang, Y. M., Chau, K. Y., Kwok, A. P. K., Zhu, T., & Ma, X. (2022). A systematic review of immersive technology applications for medical practice and education—Trends, application areas, recipients, teaching contents, evaluation methods, and performance. *Educational Research Review*, 35, 100429. <https://doi.org/10.1016/j.edurev.2021.100429>
- Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. SAGE.
- Tehrani, A. S. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: An analysis from the National Practitioner Data Bank. *BMJ Quality & Safety*, 22(8), 672–680. <https://doi.org/10.1136/bmjqs-2012-001550>
- Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 37–55. <https://doi.org/10.1037/xlm0000578>

- Theobald, K. A., Tutticci, N., Ramsbotham, J., & Johnston, S. (2021). Effectiveness of using simulation in the development of clinical reasoning in undergraduate nursing students: A systematic review. *Nurse Education in Practice*, 57, 103220. <https://doi.org/10.1016/j.nepr.2021.103220>
- Thim, T., Krarup, N. H. V., Grove, E. L., Rohde, C. V., & Løfgren, B. (2012). Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *International Journal of General Medicine*, 5, 117–121. <https://doi.org/10.2147/IJGM.S28478>
- Tracy, S. J., & Hinrichs, M. M. (2017). Big Tent Criteria for Qualitative Quality. In *The International Encyclopedia of Communication Research Methods* (pp. 1–10). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118901731.iecrm0016>
- Tun, J. K., Alinier, G., Tang, J., & Kneebone, R. L. (2015). Redefining Simulation Fidelity for Healthcare Education. *Simulation & Gaming*, 46(2), 159–174. <https://doi.org/10.1177/1046878115576103>
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>
- van Schaik, P., Flynn, D., van Wersch, A., Douglass, A., & Cann, P. (2005). Influence of Illness Script Components and Medical Practice on Medical Decision Making. *Journal of Experimental Psychology: Applied*, 11(3), 187–199. <https://doi.org/10.1037/1076-898X.11.3.187>
- Veen, M. (2021). Creative leaps in theory: The might of abduction. *Advances in Health Sciences Education*, 26(3), 1173–1183. <https://doi.org/10.1007/s10459-021-10057-8>
- Vygotsky, L. S. (1978). *Mind in Society The Development of Higher Psychological Processes*. Harvard University Press.
- Wagner, F. L., Sudacka, M., Kononowicz, A. A., Elvén, M., Durning, S. J., Hege, I., & Huwendiek, S. (2024). Current status and ongoing needs for the teaching and assessment of clinical reasoning – an international mixed-methods study from the students' and teachers' perspective. *BMC Medical Education*, 24(1), 622. <https://doi.org/10.1186/s12909-024-05518-8>
- Wason, P. C., & Evans, J. (1974). Dual processes in reasoning? *Cognition*, 3(2), 141–154. [https://doi.org/10.1016/0010-0277\(74\)90017-1](https://doi.org/10.1016/0010-0277(74)90017-1)
- Wasti, S. P., Simkhada, P., van Teijlingen, E. R., Sathian, B., & Banerjee, I. (2022). The Growing Importance of Mixed-Methods Research in Health. *Nepal Journal of Epidemiology*, 12(1), 1175–1178. <https://doi.org/10.3126/nje.v12i1.43633>
- Wheeler, D. J., Cascino, T., Sharpe, B. A., & Connor, D. M. (2017). When the Script Doesn't Fit: An Exercise in Clinical Reasoning. *Journal of General Internal Medicine*, 32(7), 836–840. <https://doi.org/10.1007/s11606-017-4018-x>
- Whittington Health NHS Trust. (2019, April 5). *Leaflet What is Sepsis?* Whittington Health NHS Trust. https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.whittington.nhs.uk%2Fdocument.ashx%3Fid%3D12951&psig=AOvVaw18INV-x7IqmuJiE3djgRzH&ust=1756854174806000&source=images&cd=vfe&opi=89978449&ved=0CBYQjRxqFwoTCNj3t_vVuI8DFQAAAAAdAAAAABAE
- WHO. (2025). *Patient safety*. <https://www.who.int/news-room/fact-sheets/detail/patient-safety>
- WordArt. (2025). *WordArt.com*. <https://wordart.com/edit/skamyr944px2>

- Wu, Y.-L., Lai, M.-C., Chang, W.-P., & Lin, Y.-H. (2025). Evaluating the effectiveness of illness script teaching on clinical reasoning skills in post-baccalaureate nursing students. *Nurse Education Today*, 144, 106401. <https://doi.org/10.1016/j.nedt.2024.106401>
- Yakhlef, A. (2010). The three facets of knowledge: A critique of the practice-based learning theory. *Research Policy*, 39(1), 39–46. <https://doi.org/10.1016/j.respol.2009.11.005>
- Yang, S., & Oh, Y. (2022). The effects of neonatal resuscitation gamification program using immersive virtual reality: A quasi-experimental study. *Nurse Education Today*, 117. <https://doi.org/10.1016/j.nedt.2022.105464>
- Yazdani, S., & Hoseini Abardeh, M. (2019). Five decades of research and theorization on clinical reasoning: A critical review. *Advances in Medical Education and Practice*, 10, 703–716. <https://doi.org/10.2147/AMEP.S213492>
- Young, M., Dory, V., Lubarsky, S., & Thomas, A. (2018). How Different Theories of Clinical Reasoning Influence Teaching and Assessment. *Academic Medicine*, 93(9), 1415. <https://doi.org/10.1097/ACM.0000000000002303>
- Young, M., Thomas, A., Gordon, D., Gruppen, L., Lubarsky, S., Rencic, J., Ballard, T., Holmboe, E., Da Silva, A., Ratcliffe, T., Schuwirth, L., & Durning, S. (2019). The terminology of clinical reasoning in health professions education: Implications and considerations. *Medical Teacher*, 41(11), 1277–1284. <https://doi.org/10.1080/0142159X.2019.1635686>
- Young, M., Thomas, A., Lubarsky, S., Ballard, T., Gordon, D., Gruppen, L., Holmboe, E., Ratcliffe, T., Rencic, J., Schuwirth, L., & Durning, S. (2018). Drawing Boundaries: The Difficulty in Defining Clinical Reasoning. *Academic Medicine*, 93(7), 990. <https://doi.org/10.1097/ACM.0000000000002142>
- Zackoff, M. W., Real, F. J., Sahay, R. D., Fei, L., Guiot, A., Lehmann, C., Tegtmeyer, K., & Klein, M. (2020). Impact of an Immersive Virtual Reality Curriculum on Medical Students' Clinical Assessment of Infants With Respiratory Distress*. *Pediatric Critical Care Medicine*, 21(5), 477. <https://doi.org/10.1097/PCC.0000000000002249>

Appendices

APPENDIX ONE – Baseline Questionnaire

Virtual Reality Simulation Study Questionnaire

* Required

1. Please Input Your Study Number *

2. Age *

- ☐ 18 - 24
- ☐ 25 - 34
- ☐ 35 - 44
- ☐ 44 - 54
- ☐ 55 - 64
- ☐ 65 +
- ☐ Prefer not to say

3. Gender *

- ☐ Male
- ☐ Female
- ☐ Non-binary
- ☐ Prefer not to say
- ☐ Other

4. Are you a medical student or qualified clinician? *

- ☐ Medical student
- ☐ Qualified clinician

5. If you are a medical student, what year of study are you? *

- ☐ Year 1
- ☐ Year 2
- ☐ Year 3
- ☐ Year 4
- ☐ Year 5
- ☐ Year 6
- ☐ N/a

6. If you are a medical student which University do you attend *

- ☐ Bristol University
- ☐ Kings College London
- ☐ University of Oxford
- ☐ N/a
- ☐ Other

7. If you are a qualified clinician, how many years of clinical experience do you have *

- ☐ 1-2 years
- ☐ 3-4 years
- ☐ 5-9 years
- ☐ 10-19 years
- ☐ > 20 years
- ☐ N/a

8. If you are a qualified clinician, what area are you specialised in? (write generalist if not specialised) *

9. Approximately how many hours of standard medical simulation experience have you had? *

- ☐ 0-4 hours
- ☐ 4-8 hours
- ☐ 8-12 hours
- ☐ 12-16 hours
- ☐ 16-24 hours
- ☐ 24-28 hours
- ☐ 28-32 hours
- ☐ 32-36 hours
- ☐ 36-40 hours
- ☐ Over 40 hours

10. Approximately how many hours of virtual reality simulation experience have you had? *

- ☐ 0-4 hours
- ☐ 4-8 hours
- ☐ 8-12 hours
- ☐ 12-16 hours
- ☐ 16-24 hours
- ☐ 24-28 hours
- ☐ 28-32 hours
- ☐ 32-36 hours
- ☐ 36-40 hours
- ☐ Over 40 hours

11. Approximately how many hours of other virtual reality experience have you had (e.g. gaming, recreational)? *

- ☐ 0-4 hours
- ☐ 4-8 hours
- ☐ 8-12 hours
- ☐ 12-16 hours
- ☐ 16-24 hours
- ☐ 24-28 hours
- ☐ 28-32 hours
- ☐ 32-36 hours
- ☐ 36-40 hours
- ☐ Over 40 hours

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

 Microsoft Forms

The figure originally presented here cannot be made freely available via ORA because of copyright.

APPENDIX THREE: Participant Information Sheet

Getting behind the headset: how can virtual reality simulations be utilised to demonstrate clinical reasoning abilities?

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: EDUC_870605

Introductory paragraph

You are being invited to take part in a research project. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

Why is this research being conducted?

Virtual Reality (VR) simulations are a new tool within medical education. Clinical reasoning is a key skill, so opportunities to develop this are crucial. A greater understanding of how VR simulations can be used to demonstrate clinical reasoning abilities would help explore the evidence for if they could be part of future teaching and assessment practices. This pilot study has the following aims:

- To determine how VR simulations can be used to demonstrate clinical reasoning abilities in medical students and experienced doctors.
- To explore whether the level of expertise of medical professionals alters the approach to clinical reasoning in a VR simulation.

Why have I been invited to take part?

I am seeking to explore how VR simulation can be utilised to demonstrate clinical reasoning across a range of clinical experience.

Undergraduate Medical Students

Qualified Doctors

Anyone who has previously participated in the Gogglesminds® VR simulations is excluded from participation in the study.

Do I have to take part?

No. It is up to you to decide whether to take part. You can withdraw from the research, without giving a reason, and without negative consequences by advising me of this decision. The deadline by which you can withdraw any information you have contributed to the research is the submission of thesis (approx. Sept 2025). If you withdraw your data before this point it will be removed from the study and deleted.

What will happen to me if I take part in the research?

You will be invited to attend a VR simulation session on site at Great Western Hospital lasting approximately 1.5 hours.

- We will rediscuss the information, I will answer any questions and if you are happy to proceed you will be asked to sign a consent form.
- You will take part in two individual VR simulations using a VR headset.
- During the first simulation you will have an opportunity to explore the environment, get used to the simulation format and practice speaking your thoughts aloud.
- Following the first simulation, you will take a short break during which refreshments will be provided and you will be asked to complete a short questionnaire regarding some baseline information (e.g. age, gender, speciality, years of clinical experience) and details concerning previous hours of simulation and VR experience.
- During the second simulation, you will be asked to complete the simulation independently whilst verbalising your thoughts.
- The visual input of the headset and audio of your thoughts will be recorded using Microsoft Teams. For clarity, you will not be visible in any recording.
- You can ask to pause or stop the research activity at any time.
- At the end of the simulation you will be asked to present the case as if referring to another clinician including your differential diagnosis.
- The researcher will complete a checklist regarding the actions you undertake.

What the simulations involve (photos of example equipment and virtual reality space below):

- 1.) You will put on a virtual reality headset.
- 2.) You will see a virtual hospital room which you can walk around. A virtual patient is present.
- 3.) You can interact with the virtual space and patient using 2 handheld controllers.
- 4.) You can talk to the patient and nurse, complete examinations, perform tests and give treatment like in the real clinical environment.
- 5.) The virtual space will update as you interact. For example, if you start oxygen, the patient's oxygen saturations will increase.

The figure originally presented here cannot be made freely available via ORA because of copyright.

What are the possible disadvantages and risks in taking part?

I have extensive experience using these VR simulations for over 12 months and rarely experience any adverse effects. However, there are some potential risks which are outlined below:

Slips, trips and falls. There is a small risk of slips, trips and falls whilst using a VR headset. To reduce the risk the research area will be a flat surface with trip hazards removed. You will be supervised at all times and the headsets have in built alerts for if you go beyond the designated safe area.

Motion sickness. Some people experience motion sickness whilst using VR headsets. To reduce the risk we will ensure that there is an adequate break between the two simulations. If you experience motion sickness we can pause immediately and stop further simulation if necessary.

Glasses discomfort. Some individuals who wear glasses find the headsets uncomfortable over glasses. If you also have contact lenses then we would recommend to consider bringing these on the day of research. If there are any challenges on the day you can pause or stop at any time.

Distress. Simulation can sometimes be psychologically distressing. Simulations utilised within this research are based on common clinical scenarios which do not include known sensitive content and you will receive a full briefing before the simulation. If you do find anything about the simulation distressing the researcher will be available for discussion during and following the session.

Identifiability. All effort will be made to ensure that your data in research outputs is not identifiable. Data will be recorded under a code number, and this will be kept separately to identifiable data. With your consent, direct quotations may be used in research outputs. It is likely you will be able to identify yourself, but it is not expected that others would be able to identify you. No data collected as part of this study will be utilised in any way within university or NHS trust assessments.

Are there any benefits in taking part?

While there are no immediate benefits for those participating in the research, it is hoped that this research will lead to increased understanding of VR simulations and more opportunities for medical education.

Expenses and payments

There will be no payment for taking part in this research. Refreshments will be provided during the break at the time of data collection.

What information will be collected and why is the collection of this information relevant for achieving the research objectives?

Your name and contact details will be collected and stored on the secure university Nexus one drive in a password protected folder for the purpose of arranging data collection and sharing the research findings with you

Consent forms will be scanned and stored on the secure university Nexus one drive in a password protected folder – following scanning the original paper copy will be destroyed.

Questionnaire regarding basic demographic data, years of clinical experience and data concerning previous exposure to simulation and VR. I am interested to discover how individuals at different levels of clinical experience approach the simulations and so this information will help ensure that I am gathering data from a range of individuals with differing expertise levels. The data will be collected using Microsoft forms during the research session under an anonymised code number.

The second simulation will be recorded in the following ways

1. An audio recording via Microsoft teams to record your verbalised thoughts and reasoning.

2. A recording of the video INPUT of the VR headset. You will not be visible in any recording.
3. The researcher will complete a checklist regarding the steps taken in the simulation.

The recorded data will be stored on a university Microsoft Teams account and then downloaded to the secure university Nexus one drive in a password protected folder. The audio recording will be transcribed and the video input feed of the headset analysed.

Transcriptions, video analysis records and the checklist will be stored under a code number separate to any identifiable data on the university Nexus One drive under password protection.

Code number: A document linking the anonymised code number to your name will be kept separately to other data in the secure university nexus one drive in a password protected folder. This is to enable removal of your data should you wish to withdraw your data before the submission of thesis (approx. Sept 2025). ONLY the researcher and my research supervisor will have access to this.

Only those with a legitimate reason will be able to access the data, this is the research team and IT/technical support.

Data will be maintained for the following lengths of time:

- Consent forms: 3 years
- Name and contact details: For a maximum of one month after release of research output to enable this to be shared with yourself.
- Audio and headset visual input feed recordings: For a maximum of one month after thesis submission.
- Document linking the code number with your name: For a maximum of one month after thesis submission.
- Any data stored under your code number will be stored for 3 years after publication or public release of the work of the research.

As this is an initial study the data recorded under an anonymised code number (without personal identifying information) may be available to future researchers on request to help further develop this research.

Will the research be published? Could I be identified from any publications or other research outputs?

The findings from the research will be written up in a masters thesis and may be published in a peer reviewed journal and conference presentations/ posters.

I would like your permission to use direct quotations but without identifying you in any research outputs.

A copy of my thesis/ dissertation will be deposited online in the [Oxford University Research Archive](#) where it will be publicly available to facilitate its use in future research.

Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance website at <https://compliance.admin.ox.ac.uk/individual-rights>.

Who has reviewed this research?

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: EDUC_870605).

Who do I contact if I have a concern about the research or I wish to complain?

If you have a concern about any aspect of this research, please contact Dr [REDACTED] or Dr [REDACTED] and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the University of Oxford Research Governance, Ethics & Assurance (RGEA) team at rgea.complaints@admin.ox.ac.uk or on +44 (0)1865 616480.

Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Dr [REDACTED]
Department of Education
15 Norham Gardens
Oxford
OX 6PY
University tel: 01865 274 024
University email: [REDACTED]

Department of Education
15 Norham Gardens,
Oxford, OX2 6PY



APPENDIX FOUR: Consent Form

Consent to take part in: Getting behind the headset: how can virtual reality simulations be utilised to demonstrate clinical reasoning abilities?

Central University Research Ethics Committee (CUREC) approval reference: **EDUC_870605**

Purpose of Study: A pilot study investigating how virtual reality simulations can be utilised to demonstrate clinical reasoning abilities and explore any differences in approach depending on experience.

**Please initial
each box if you
agree with the
statement**

I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

☐

I understand that my participation is voluntary and that I am free to withdraw at any point until submission of the research output (01/09/2025) without giving any reason.

☐

I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.

☐

I understand that I will not be identifiable from any publications, thesis or other research outputs including conference posters/presentations.

☐

I consent to being audio recorded.

☐

I consent to the virtual reality headset visual input being recorded.

☐

I understand how audio and visual input recordings will be used in research outputs.

☐

Use of quotations: Please indicate your preference (select *one* option):

a) I do not wish to be quoted. **or**

b) I agree to the use of quotations in research outputs if I am not identifiable.

☐
☐

I give permission for you to contact me again to clarify information and share research outcomes.

☐

I understand how to raise a concern or make a complaint.

☐

I agree to take part. I hereby assign to the researcher all copyright in my contribution for use in all work stemming from the project and future projects.

☐

 Name of participant

dd / mm / yyyy

Date

Signature

 Name of person taking
 consent

dd / mm / yyyy

Date

Signature

APPENDIX FIVE: Model Answer Checklist

Adapted from Sepsis Guidelines NICE 2024

1.3 Initial assessment and examination										
Physical Observations - NICE 1.3.1						NICE 1.3.7	NICE 1.3.10			
Temperature ☐	Heart Rate ☐	RR ☐	Oxygen Saturation ☐	Blood Pressure ☐	Level of Consciousness ☐	Urine Output ☐	Clinical Examination ☐			
1.13 Managing suspected sepsis in acute hospital settings										
NICE 1.13.1		Bloods Tests NICE 1.13.2							NICE 1.13.3	NICE 1.13.4
Bloods samples pre antibiotics ☐	Microbial samples pre antibiotics ☐	Blood Gas (inc lactate and blood glucose) ☐	Blood Culture ☐	Full Blood Count ☐	Urea and Electrolytes ☐	Liver Function Tests ☐	CRP ☐	Clotting Screen ☐	IV Broad Spectrum Antibiotics ☐	IV fluids ☐
Antibiotic therapy, intravenous fluid and oxygen 1.14 1.15 and 1.16										
NICE 1.14.2	NICE 1.15.8	NICE 1.16.1								
use existing local antimicrobial guidance ☐	Reassess after fluids and give second bolus if indicated ☐	Oxygen to target saturations 94% to 98% ☐								
Finding the source of infection 1.17										
NICE 1.17.2										
Urine Culture ☐	Sputum Culture ☐	CXR ☐								

APPENDIX SIX: Categorisation of Activities					
Category	Activities				
Asking Patients Medical History	Using drop down menu of questions	Listening to answers			
Examining patient	Checking physical observations (respiratory rate, oxygen saturations, heart rate, blood pressure, temperature, capillary refill rate)	Checking blood glucose	Auscultating for lung sounds	Auscultating for heart sounds	Other verbalised examination checks (e.g. looking for swelling)
Prescribing Treatment	Ordering intravenous fluids	Ordering medication			
Giving Treatment	Giving intravenous fluids	Giving antibiotics	Placing urinary catheter		
Ordering Investigations	Ordering Xrays	Ordering ECG	Ordering blood tests	Ordering culture samples	
Reviewing Investigations	Reviewing Xrays	Reviewing ECG	Reviewing blood test	Reviewing cultures	
Using Medical Guidelines	Using decision aids (CURB65)	Using medical guidelines	Using antibiotic guidelines		
Handing Over to another Clinician	Verbally handing over				

APPENDIX SEVEN: Codebook

Code	Meaning	Example
Hypothesis generation	Initial or subsequent statement of a hypothesis	"Just talking to him. Right, that sounds like you've got an infection"
Interpreting finding linked to hypothesis	Linking findings to an already determined hypothesis to increase/ decrease suspicion	"Um, CRP is high. Ok, so it's in keeping with the infection"
Justification statements	Statements used to justify actions	"I would like to know what his kidneys are doing because I'm worried about sepsis, so urea and electrolytes would be reasonable"
Linking evidence together	Statements which link together two or more pieces of information	"The chest Xray ... Tracks with what we heard on his chest"
Diagnostic conditional Reasoning	Any if/then conditional reasoning where x finding = y action related to diagnosis	"I can hear coarse crackles, and his saturations are quite low, so I'm going to order a chest Xray"
Management conditional Reasoning	Any if/then conditional reasoning where x finding = y action related to management	"So, he's hypotensive so let's give you a bag of fluid"
Problem representation	Collating and summarising findings and actions to re frame the problem	"Um, so I'm thinking he's short of breath, he's got a cough, he's maybe a little bit feverish, a bit confused. So, I'm thinking about a pneumonia"
More complex schema	Examples of more complex conditional reasoning (if/then/and or if/then/unless)	"So, let's pop some oxygen on. Before we do that, let's just make sure he's not got any COPD"
Less developed schema	Examples of less complex conditional reasoning	"Um, he's got a cough as well, but no temperature, so that's quite odd. So probably not a pneumonia"
Evolving schema	Examples where schema may be developed with addition of new information	"White blood cells are high. Ok, so it, it seems like there is an infection, he's just not febrile"
Unverified assertion	Statements made not backed up by any evidence or factually incorrect within the scenario	"He is a male over the age of 65 who had a background of a fever" (no fever)
Identifying context	Statements highlighting contextual factors which impact thinking	"So, I see an older gentleman in front of me"

Interpreting finding	Statements interpreting a finding but without expressing clear challenge or confidence	"Right, okay. Looks like he's got some, definitely got some consolidation round the. What sides that? The right, right base"
Quantifying goal	Statements setting particularly targets or goals of actions	"We'll aim sats 94 and above. We can wean that down as it's necessary"
Triangulating data sources	Triangulating one source of data with another	"don't think he has any medication or any past medical history, let's just corroborate that with, with his summary care record"
Weighing up/ considering	Statements where evidence is considered before reaching conclusion	"So, he said he was not passing much urine, so I'd maybe like to do a catheter to measure his output. We could do this sorta, non-invasive. But it's a shame, but he's very confused, so he might not tell the nurses when he's going to the toilet. So maybe we'll just do a catheter"
Reassess	Review of actions and effects	"Ok, so that saturation is coming up nicely, his heart rate is gone down a little bit since he came in"
Situating statements	Statements that indicate position/ progress in the flow through scenario	"Ok, so I'm just gonna set up some basic observations to start off with."
Use of Scaffolding	Where available information is used to help structure actions and decisions	"Ok. So, we've got a CURB score of three. So, I'm going to go back to the guidelines. I think this is severe"
Use of A-E structure	Any instance of explicit or implicit use of the A-E structure	" Ok, so I'm happy with the airway, so I'm gonna move on to breathing."
Check listing	Use of a checklist structure to review or plan actions including sepsis 6	"We've given oxygen. Sorry, we've taken lactate, we've taken cultures. Take. We need a urine output"
Confidence in interpretation	Statements which indicate succinct or confident investigation interpretation	"Okay, chest Xray. It shows some, we've got a right lower lobe pneumonia"
Expressing uncertainty	Words or statements that convey degree of confidence in statement made	"Um, <u>maybe</u> at the bases <u>maybe</u> some kind of pneumonia"
Expression of concern	Words or statements that convey concern or highlight worry about an item	"Okay, I'm <u>slightly worried</u> about his oxygen"
Reflection in action	Statements that indicate reflection on own actions or thoughts	"Asked allergies - Should have asked that before I gave him a penicillin, but there we go, we live and learn"

Expression of non-concern	Words or statements that convey non-concern or contentment about an item	"Um. So, the oxygen sats are coming up. That is good"
Patient centred approach	Words speaking directly to patient or considering them/ family	"I'm sorry (to patient talking about wife passing away)"
Decision statement	Statements made to indicate a decision-making threshold has been reached	"So, I think I'll get some antibiotics into this patient. Sounds like it could be a pneumonia so. He didn't have any allergies. So, amoxicillin"
Expressing confidence	Words or statements that convey degree of confidence in statement made	"we'll <u>do definitely</u> a blood culture, lactate, sugar and U's and E's to start with"
Interpretation Challenge	Any challenges encountered in interpreting an investigation	"ECG. Ok. So. This isn't a normal, it's regular, but there aren't any. There are P waves. They don't seem to be there (looking at another lead) so maybe there are no P waves."
Calculation challenge	Challenges with accurately calculating scoring systems	"Uh, he's not confused, I think. And the urea? Oh, wait, urea is in the blood testing. Look for the urea. That was 28 (highlighting RR) Done a blood pressure. Less than 90? Not anymore. So, he'd be a 1 or a 2, I'll check the urea. So, urea more than 7, he'd be a 2. Ok, so he's a CURB 1."
Inferring from less direct data	Conclusions are drawn from logically inferring rather than direct data/ observation	"Ok, wandering round and getting lost, that doesn't sound very good, but could indicate confusion"
Misinterpretation	Misinterpretation of objective data	"I wanted the urea didn't I to ... So, it is slightly raised, not high enough to score though" (did score)
Incorrect inferences	Incorrection conclusions are drawn from logically inferring rather than direct data/ observation	"So, he's not confused"
Technology issues	Situations where the technology did not work as intended	"Participant clicks to give blanket. Researcher: Oh, I don't know ... Well, we will say it has worked" (not working)
Cognitive overload	Statements or part sentences which may indicate cognitive overload	"I've lost track sort of what I've ordered and where it's going to, what the um, where the results are"

Technology limitations	Situations where actions wanting to be undertaking are not possible under sim	"Researcher - There isn't a way to do those at the moment, just kind of how his pupils look from what you can see"
Explaining technology	Situations where researcher needs to explain how to complete a task within the sim	"Researcher - yeah. So, if you take the stethoscope in your hand. Now go behind him with the stethoscope. There you go"
Fidelity	Where the technology is unable to reflect reality in a way that impacts the process/ approach	"So, it looks like there's consolidation here (pointing at right lower zone of lung). Which wouldn't really match with the crackles up here (pointing at patients right upper zone), But I guess that's a code thing?"
Over scaffolding	Where functions specific to the VR environments are utilised in a way to provide reasoning/ diagnostic/ management hints in a way not possible in the genuine clinical environment	"Checks through clinical tasks menu"
Scaffolding decisions	Use of guidelines/ a senior to help scaffold a decision	"So, don't know about this. Is there any way of getting senior help?"
Experiential knowledge modulating action	Situations where experiential knowledge effects decision making or expectations	"Okay, I think it might be worth getting another cannula in because fluid running as well"
Lack of experiential knowledge or uncertainty with sim metrics	Situations where it is hard to know if decisions are related to gaps in experiential knowledge or uncertainty with how sim works	"Um, I might repeat the bloods. Now he's had some fluid... Oh so they don't change when you repeat them"

APPENDIX EIGHT: Example Researcher Notes

- Approx 20 mins to introduce, sign consent and trial practice sim (in combo with another participant at both participants request)
- Approx 20 mins break and questionnaire (headset charging time)
- Approx 20 mins to complete recorded simulation -individual.
- Approx 10 mins to debrief – individual.

Challenges

- Poor audio pick up
- Initial quiet period and then more verbalising after initial prompt
- Several areas needing researcher intervention to find or enact plan
- Headset battery issues

Immediate thoughts

- Quick early mention of sepsis linked to confusion
- Then concluded unsure of current evidence of sepsis in handover - ? how to code this. However, others do not mention in top differential even if mentioned earlier
- Focus on history completion, then A-E
- Used guidelines and selected correct antibiotics but did not use CURB
- CXR prior to auscultation
- On asking after – for them main clue for sepsis was confusion
- Participant did note that short number of guidelines available may be leading
- Participant did note that VR patient looks and acts far more well than observations suggest

Changes

- Spare battery for headset

APPENDIX NINE: Example Reflexivity Log During Analysis

Date: 25/07/2025

Dataset/Transcript: First full readthrough

1. Initial Thoughts

- Seems to be different overall pathways taken
- 1.) ½ q's – A-E – fuller history
- 2.) Full history – A-E
- 3.) ½ Q's – A-E
- Huge volume of data is descriptive ? link to cognitive load
- If/then reasoning demonstrated in all transcripts
- The most junior and the two most senior participants did not explicitly utilise an A-E structure
- Use of ? memory aid (need to consider correct term) due as A-E and sepsis 6 clearly evident in early career
- Demonstration of certainty - ? related to language (junior = lots of maybe, longer descriptors rather than pinning to diagnosis, seniors more confident language)
- Some form of differential and then testing seen in all
- Integration of action ? differentiate management/ diagnostic reasoning
- What is the decision threshold
- Early suggestion of sepsis in most senior
- What is the role of knowledge?
- Reflection in action? Role of metacognition
- Possibly not relevant to research Q but very interesting no one remembers the age
- Errors of commission (e.g. ordering trop) errors of omission (not identifying sepsis) and just plain errors (very poor rates of correct CURB calculation)

2. Coding Decisions

New codes:

Not formally coded as familiarisation

Codes changed:

Areas of uncertainty:

3. Interpretive Reflections

- Until reading transcripts, I have under recognised the role of technology challenges. Need to be cognisant of if need to intervene to explain sim mechanics may influence data
- May own stage and training is likely more in keeping with early career – am I more able to recognise reasoning in this group as it aligns more with my own?

4. Evolution of understanding

- First full read through – no codes/ potential themes to develop understanding of yet.

5. Relationship to narratives

- Awareness (either as junior or senior) to most participants and impact this may have had
- Recognition of potential feelings of professional jeopardy in participants, especially seniors and the impact this may have
- Some narratives are less descriptive and more link actions to thinking – although these may lead to greater ability to interpret it important to give more even weight

6. Developing analysis

- A range of reasoning strategies appear to be demonstrated with consistency across different levels of expertise
- Seems to be greater confidence in language around interpreting findings/ representing problems with greater experience. ? decision thresholds
- Sepsis appears in the differential of all apart from the 2 most novice participants (but not in the final handover for all)
- Various flow through the sim, but not consistent with expertise

7. To Revisit

- Need to relook at timings for urine output – change to when orders catheter/ mention fluid output to make more consistent
- Record first mention of chest infection/ pneumonia, may be interesting
- Quan data may suggest some role of pattern recognition but is this evident in transcript?

APPENDIX TEN: Ethical Approval Letter



Education (Educ) DREC
15 Norham Gardens, Oxford, OX2 6PY

Applicant: [REDACTED]

Principal Investigator: [REDACTED]

Department: Education

Study title: Getting behind the headset: how can virtual reality simulations be utilised to demonstrate clinical reasoning abilities?

(Version: 1.0)

Ethics reference: Education (Educ) DREC - 870605

Dear [REDACTED]

On behalf of the Committee, I confirm that the above research study described in the application and other supporting documentation submitted to the committee has been carefully considered by the Education (Educ) DREC in accordance with the University's regulations and policy for ethics approval of research involving human participants, human tissue and/or personal data. The opinion is as follows:

Opinion of Research Ethics Committee: Favourable Opinion

Subject to the following conditions:

Decision Date: 2 Jan 2025, 17:25

Opinion End Date: 2 Jul 2026

If favourable, insurance-provided indemnity arrangements will be in place between the decision date and opinion end date and you may now commence your study activities. Should you plan to continue the research beyond the end date above, it is your responsibility to ensure that you request, and receive, an extension (via amendment) from the committee for indemnity to remain in place. You may be required to provide a justification.

Please note the following:

Amendments: Should there be any subsequent changes to the reviewed study, applications for amendments can be made via the Oxford Ethics Application System (Worktribe Ethics).

Reports: Studies considered by OxtREC are expected to submit an *annual progress report* on each anniversary of study approval, until the study is completed. An end of study report is also required.

Audit: This study may be selected for audit at the discretion of the Research Governance, Ethics and Assurance Team.

Data safety: It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely in accordance with University requirements. Further guidance and advice are available from the [Research Data Team](#).

Additional information is available at <https://researchsupport.web.ox.ac.uk/governance/ethics>

Yours Sincerely

Education Ethics Officer

APPENDIX ELEVEN: Local Approval Letter



Swindon Academy

Undergraduate Office
Great Western Hospital
Marlborough Road
Swindon SN3 6BB

19/11/2024

Dear Dr [REDACTED]

The ethical committee approves your medical education research at Great Western Hospital and research related to your Master's degree in medical education. Any additional IRB approval from an external university will also be satisfactory for any current or future projects related to their students at our site for educational purposes.

Yours sincerely,



Dr Chris Jacobs
Chair

SAMER
Swindon Academy Medical Education Research Committee

The Academy
Great Western Hospital
Marlborough Road
Swindon
SN3 6BB

APPENDIX TWELVE: Risks and Mitigations

- 1.) Slips, trips and falls. Likelihood – Unlikely. The environment will be chosen for its suitable for carrying out virtual reality simulations including a flat floor surface. The headsets have inbuilt safeguards to ensure users to not travel beyond pre-set boundaries and users will be supervised at all times during use. In the unlikely event of any slips, trips or falls a first aid kit is present at the research site, I am a medical professional and will be collecting all data, and the research area is on a hospital site.
- 2.) Motion sickness. Likelihood – Possible. Motion sickness (or cyber sickness) is known to be experienced by some individuals using this technology. A midpoint break is included in the research protocol to reduce the likelihood of experiencing these symptoms. Should motion sickness occur the research will be immediately paused and stopped if necessary. Any symptoms are known to typically resolve quickly after stopping using the device.
- 3.) Glasses discomfort: Likelihood – Possible. Some individuals who wear glasses find the headsets uncomfortable over glasses. If participants also have contact lenses, I will recommend that they consider bringing these on the day of the research. If there are any challenges during data collection the research can be paused or stopped at any time.
- 4.) Psychological safety: Likelihood – Unlikely. Participants might experience psychological distress during virtual reality simulation session. The content of the virtual reality simulations has been chosen to be a common clinical scenario without sensitive content. A briefing will be given before commencing the scenario to ensure that participants are aware they can pause or stop at any time, and the researcher will be available to discuss further if any issues arise.
- 5.) Participant identification. Likelihood – Rare. Due to the relatively small number of participants and the challenges of fully anonymising qualitative data there is a small chance that individuals will be identifiable from their contributions.

-All research data will be recorded under a pseudonym, with this data being kept separate to personal identifiable data.

- Information regarding demographics and medical speciality will be presented collectively for the participant groups to reduce chances of identification.

- Participants will be advised in the information sheet that their data will be pseudonymised, rather than anonymised in the research. It will be advised that they will likely be able to identify their own contributions from research outputs, but it is not expected that others would be able to identify them.

APPENDIX THIRTEEN: Recruitment Flow Sheet

