A mixed methods study to explore the process of clinical reasoning employed by clinical students during primary clinical assessments

Dr James Gray

School of Medicine and Population Health

The University of Sheffield

Submitted for PhD

07/08/2024

Acknowledgements

I would like to express my sincere thanks to my supervisors Professor Michelle Marshall and Dr Joanne Thompson for their support, wisdom and encouragement through this PhD.

I would also like to thank Jade Moore, Amy Donaldson-Perrott, Dr Dyfrig Hughes and Professor Julian Burton for their support in allowing me to engage with the various student bodies and supporting the data collection.

To all the students who participated my huge thanks, especially those who gave up their time to take part in the focus groups, and the support from the Sheffield Medical Student Society.

Finally thanks to my wife Clare for her seemingly unlimited patience and tolerance during this process. I could not have done it without you.

Abstract

Background and Purpose

Clinical reasoning is required at all points of patient interaction yet is confused due to a lack of coherence in the language used and description within the literature. Clinical educators must develop the skills of clinical reasoning amongst students as a core skill of the professions. Failure to do so risks students having good knowledge but a lack of effective cognitive tools to apply it. Understanding how students undertake the clinical reasoning process as they leave clinical programmes can enable us to ensure the appropriate development of educational interventions to develop their expertise in this skill.

This research project sought to answer the question as to how clinical students undertake the process of clinical reasoning in Single Best Answer (SBA) questions and Objective Structured Clinical Examinations (OSCE).

Methods

A narrative literature review was undertaken to formulate a concept framework for the clinical reasoning process and explore the implications of the literature regarding education and assessment. From this a working definition for clinical reasoning was formulated for use in this study.

The concept framework was used as the basis for a nominal group approach to formulate ten key questions to use in the primary study to explore different aspects of clinical reasoning as set out in the concept framework.

The primary study was a convergent mixed methods study. The questions generated from the nominal group were used in the quantitative arm, as questionnaire items, and the qualitative focus groups as the interview guide. In the quantitative arm students answered 40 SBA questions, or completed an OSCE, and then immediately answered the ten-item questionnaire. In the qualitative arm students were interviewed in focus groups. A pilot of the methodology was conducted to ensure feasibility.

Results

Three universities took part in the study, two with both medical and physician associate students and one with physician associate students.

In the quantitative element a total of 595 data points (an individual participant completing all 10 questions) were collected, 291 for SBA and 304 for OSCE. Statistical analysis was undertaken in Microsoft Excel and SPSS (IBM, 2021; Microsoft, 2023). The quantitative results permitted comparison of groups with different characteristics. These showed differences in reasoning between SBA and OSCE, but these were very small. There were three questions with statistically significant differences, in both SBA and OSCE, between medical and PA students although these were small. There was a slightly higher tendency toward pattern recognition in medical students. There was one question showing a statistically significant difference between genders for both SBA and OSCE. Observed difference was small however females felt being in an assessment affected their reasoning more. There was one question showing a statistically significant difference between ethnicities for SBA and three for OSCE. The observed difference was, again, small.

The qualitative arm included seven focus groups of between 2 and 18 participants and a total of 18 medical students and 28 physician associate students. These were transcribed and analysed using thematic analysis (Clarke and Braun, 2013). Four key themes were identified. The qualitative results and quantitative results were then converged, and conclusions drawn.

- Students felt being in an exam affected their reasoning, but this appeared linked to the difference between eustress and distress.
- SBA is heavily pattern recognition, OSCE less so than the quantitative data alone would suggest. OSCE discrepancy (e.g. patient appearance differing from the instructions) lowered the value of demographic information and clinical setting in the stem.
- Early hypothesis generation was typical in both assessments, but SBA was answer driven, OSCE was plan driven.
- Student thinking process tended toward inductive reasoning, but this was stronger for SBA.
- Evidence for metacognition was seen. The results suggest abductive reasoning is the most appropriate description which is about using the data available to draw a “most likely” conclusion.
- There is some evidence for students getting to answers without really understanding the cognitive approach they have taken however this appears to link to satisficing in a position of lack of knowledge rather than being linked to inductive reasoning.
- Satisficing occurs commonly but is multifactorial in nature, particularly around knowledge and time pressure.
- Students are aware of biases, although there is confusion with heuristics. The evidence that they actively seek to manage them is less clear.

Discussion and Implications

The results facilitated a revision of the concept framework and consideration of number of key areas. Issues for assessment include time pressure, question ordering, assessment schemes and implications for workplace-based assessment. Issues for education include diagnostic stewardship, situated cognition, teaching approach, bias and technological implications.

The convergent results show a number of implications for education, policy and research.

For education improving clinical reasoning will require a shift in the way that curricula are planned and delivered. This will have implications for faculty training, including in the workplace with our non-core staff, particularly more junior clinicians. This work also challenges our assumptions that our assessments are effective in ensuring that students are appropriately prepared for decision making in practice.

For policy, the equal footing of clinical reasoning with respect to clinical knowledge needs to be emphasised by regulators and educational bodies. Evidence demonstrates that clinical reasoning errors directly lead to poor patient outcomes thus instilling a requirement for this in policy makes sense.

Further research on aspects of clinical reasoning within assessments and curricula is crucial including looking at workplace-based assessment and whether it compliments these current methodologies.

Further research into the delivery of education in clinical reasoning, both in terms of faculty and student understanding, will further improve the delivery of education in this area.

The design, and subsequent revision, of the concept framework through this piece of work can act as a guide for educators when designing programmes and underpin the education and curricula plans.

Contents

List of Tables and Figures

Tables

Figures

Chapter 1 – Introduction

1.1 The cognitive imperative for clinicians.

It is essential for clinicians to have extensive knowledge in order to practise, however the knowledge alone is insufficient if it cannot be applied to produce a safe and effective diagnostic or therapeutic endpoint. This cognitive imperative is crucial if we are to produce fully rounded, and safe, clinicians. During the patient journey a clinician may need to undertake a number of these cognitive processes with multiple diagnoses and therapeutic options being considered. This is something that is often poorly reflected within the literature.

Failure to apply the knowledge effectively will lead to a risk of misdiagnosis, or delivery of inappropriate therapies, with the risks that both of these bring. Misdiagnosis has previously been considered to cause significant patient risk in practice (Saber Tehrani *et al.*, 2013) whilst medication errors, as a proxy for cognitive errors in therapy management, also lead to significant patient harm and economic burden within healthcare (Elliott *et al.,* 2021).

Whilst in medical education the assessment of these cognitive processes is most commonly undertaken through assessments based upon clinical scenarios, or vignettes, with an expectation that these will drive a similar pattern of thinking to that which will occur in the clinical environment. Single best answer (SBA) and objective structured clinical examinations (OSCE) remain the mainstay of these assessments.

Clinical educators are required to assist students that they work with to develop their cognitive skills as clinical care is, in essence, knowledge application within ever changing clinical cases. It is likely that educators will tend toward promoting cognitive processes with which they are most familiar through their own experience, positive or negative. This has the potential to stifle the development of a rounded education in the subject, thus risking the production of clinicians who may be limited in their ability to apply their learning effectively for patient benefit. It is important that clinical educators understand how students undertake these processes in order to ensure the appropriate development of educational interventions.

One of the most significant challenges with this application of knowledge is the multitude of terms utilised by practitioners and researchers when discussing it, such as 'clinical reasoning' and 'clinical decision making', with a consequent lack of coherence to its study, nature and outcomes. When considering, and researching, this process the medical profession has traditionally focussed on diagnosis, taking a position that reduction in diagnostic errors can lead to improved patient outcomes (Gruppen *et al.*, 1988; Graber *et al.*, 2002). The psychological professions and therapy professions (particularly occupational and physical therapy) have tended to focus more on the process of patient management and therapeutics (Unsworth, 2004; Chaffey *et al.*, 2010; Andrews and Syeda, 2017). These two different focuses have led to variances in conceptual modelling and descriptions of the process. Within the nursing profession nurse practitioner descriptions have tended to focus toward the medical concepts of this process (Baid, 2006) whilst nurses working in the ward environment have tended to align more toward the decisions by which treatment and patient management decisions are made (Holder, 2018).

The other key challenge is that of context. Clinicians work with patients within a specific context or the "background noise" that relates to, and influences, the cognitive process, whilst not being truly part of it. Creating a research situation in which a single context is separated out is challenging given that the concept of context is widely defined and sits on a wide spectrum from the physical location in which a decision is being made through to the emotional state of the parties involved.

1.2 The Researcher

I am an experienced academic working within the School of Medicine and Population Health as course director for the Physician Associate MMedSci, director of student recruitment for the school, and contribute to supporting the Medicine MB ChB through teaching and across all assessment processes. I continue to practice clinically as a General Practitioner alongside my academic role. My background includes being part of the development group for a new examination at the Royal College of Surgeons of Edinburgh (Diploma in Urgent Medical Care), a regular examiner for the Physician Associate National Examination and an experienced external examiner for both Physician Associate Studies and Medicine.

Due to the above there are considerations regarding my relationship with the study participants, particularly around the concept of the insider researcher and power gradients. This is discussed in detail in chapter 11, where limitations and mitigations are discussed, along with the way these have been addressed.

1.3 Purpose of The Study

This study will explore clinical reasoning utilised by students when undertaking SBA and OSCE assessments. The term clinical reasoning is widely used in medical education to describe the cognitive processes of the application of knowledge and these examinations remain the mainstay of clinical education. Understanding how students are applying their knowledge in this specific context can allow us to better understand how our education is influencing students, or how it needs to be adapted. This can inform us, enabling us to better provide students with the necessary tools to apply their knowledge across the patient journey. The study target group is final year medical, and physician associate, students.

1.4 The Process

As stated above there is a lack of conceptual clarity regarding clinical reasoning. Gaining clarity regarding this term is essential given its routine use within medical education. An initial narrative review, chapter 2, will inform the development of a concept framework of the nature of this cognitive process, facilitating a clear operational definition for clinical reasoning to use through this study, and with the potential for wider application. The research question will be developed through this review and set out at the end of chapter 2. A consensus group method will then employ the concept framework to develop the questions that will form the basis of a mixed methods study. The study data will be gathered contemporaneously during assessment (quantitative) and through focus group discussions (qualitative) with a convergent approach to the data analysis. A narrative conclusion will enable the findings to be explored, reflected upon, and the conceptual framework refined.

The findings are designed to better inform clinical educators and facilitate curriculum design that ensures that the application of knowledge can form a golden thread through both medical, and physician associate, course teaching and assessment informed by the preferred approach as it is now and the gap to where we would like it to be.

Chapter 2 - Narrative Review – Conceptualising Clinical reasoning

2.1 Introduction and Chapter Summary

This chapter is a narrative review designed to develop a concept framework of the nature of clinical reasoning. It will initially discuss the background to the choice of review type and the search strategy employed. It will then explore the following areas:

1. Concepts of reasoning – primary reasoning models and their underpinning concepts
2. Alternative theories and concepts of clinical reasoning – this section explores diagnostic reasoning in more depth, including the finalising process and risks to it. It will also look at differences in therapeutic reasoning.
3. Proposing a concept framework for clinical reasoning – this section will bring the discussed concepts together to propose a concept framework for clinical reasoning that will underpin the research. This section will include some illustrative case studies.
4. Implications for Clinical Education – this section will look at the underpinning theory of the teaching of clinical reasoning in more detail.
5. Assessment of reasoning – this section will consider the approaches and challenges to the assessment of clinical reasoning with particular reference to the primary assessment types used in clinical education.

For the purposes of this narrative review the term decision-making will be used whilst the concept framework is constructed. From this an operational definition of clinical reasoning, to be used in this study, will be produced to aid the reader. After consideration of the implications for education and assessment a short conclusion will set out the research question to be answered.

2.2 Undertaking a Review on this topic

Undertaking a literature review focussed on the topic of decision-making within the clinical context presents a conceptual, and source material, challenge. The use of multiple interchangeable terms such as “critical reasoning”, “diagnostic reasoning”, “critical

thinking", amongst many others, combined with much of the literature being theoretical and conceptual, makes the development of a coherent, and definitive, concept framework challenging. Authors have noted that healthcare professionals work in an area without absolute right and wrong, with research into this area therefore not amenable to the "gold-standard" of randomised controlled trials and thus concepts of this process, in all its forms, must be seen through this lens (Higgs *et al.*, 2001).

The variance in research between the clinical professions is partially what led to a cross professions study being undertaken in 2018 using a "pause and reflect" approach. Multi-professional experts sought to produce a unified approach to conceptualising this decision-making process but struggled to do so, leading to the conclusion that "it is unlikely that a unified definition of clinical reasoning is achievable or even desirable" (Young *et al.*, 2018). The authors comment that research into clinical reasoning tends to be undertaken within silos, with each researcher championing their own personal theories, putting up boundaries to awareness of other concepts, and consequently each group puts limitations on its thinking (Young *et al.*, 2018). This also leads to the requirement for any work in this area to be explicit regarding the operational definition of the term clinical reasoning that is to be used. In recognising that this decision-making is not exclusive to any single profession or level of practitioner it is notable that a structured review of this process within the nursing literature found that a higher level of education does not necessarily correspond to better decision-making skills. The authors found that experience and informal knowledge were as important to the decision-making process as more formal knowledge (Simmons, 2010). It can be deduced that all professions, levels of practitioner, and parts of the clinical encounter should be considered when conceptualising the process if we are to ensure a model that can be utilised by all. In exploring the different cognitive models, and descriptive concepts, in this way there is opportunity to draw them together to form a more inclusive concept framework, which aims to work across the full clinical encounter. It is important to recognise that any concept must stand up to scrutiny against decision-making both in the formulation of diagnosis, or differential diagnosis, and the management plan that stems from it. Practitioners may undertake one, or more commonly both, of these decision-making processes depending upon circumstances of any specific clinical interaction.

The comments of Young et al might make it appear an act of futility to seek to produce such a concept framework, however their statement relates to the issues of poorly defined contexts within which much of the research is set, making generalisability from the literature challenging. Through this narrative review of the literature, I develop a concept framework, utilising an effective selection of literature, from which an operational definition of clinical reasoning can be defined for use within this study and allow improved clarity, and thus generalisability, of the outcomes in relation to the research question.

2.3 The Narrative Review

A narrative review is different to a systematic review in that it is designed to give a broad, and easily digested view of a topic through a selection of the literature, rather than address a clear and tightly defined topic in exhaustive detail. The narrative review looks at topics with more breadth, and with multiple study designs, seeking to distil the key information from appropriate sources without the same level of scrutiny around study design (Murphy, 2012). It is noted that there are no formal guidelines for the way a narrative review should be written (Gregory and Denniss, 2018).

This topic is more appropriate for a narrative review than a systematic review due to the nature of its breadth, along with the plurality of study types, commentary papers, and opinion pieces that form much of the literature. Pragmatic decision-making in the choice of search terms, and approach to managing the evidence base, will enable a sufficient breadth of literature to be extracted, explored and appropriate inferences to be drawn. From this it will be possible to develop the theoretical concept framework, a clear statement of an operational definition of clinical reasoning to be made and allow an exploration of this process in relation to clinical education and assessment.

Whilst there are no formal guidelines as to the way such a narrative review should be constructed, it is important to address its quality. In order to ensure that the review was of an appropriate quality, a benchmarking process was undertaken against the SANRA scale for measuring the quality of narrative reviews. (Baethge *et al.*, 2019). The draft narrative review was then reviewed against this scale by an independent colleague within the School of Medicine and Population Health as a further check to ensure quality.

2.4 Search Strategy

Interchangeable terms such as “clinical reasoning”, “clinical judgement” or “critical thinking” led to difficulty in defining suitable search parameters. Including two or more of these terms would routinely lead to over 1000 references being retrieved leading to a process of trying different approaches to obtain an appropriate cross section of literature. This issue has been noted by previous authors in this area (Simmons, 2010).

A literature search was undertaken utilising Web of Science, SCOPUS and EBSCO to cover all key databases in medicine, the wider therapeutic professions and education. The search was defined using the Boolean operators “clinical reasoning” AND “definition” OR “theory of”. The search was limited to the English language. No specific time periods for the literature were set however some databases are date limited as stated in the table. The process is detailed in Appendix 1 and database coverage in table 1.

Table 1: Relevant Database coverage within Web of Science and EBSCO

Web of Science	EBSCO
Web of Science Core Collection, 1900-present	British Education Index
BIOSIS Citation Index, 1926-present	Child Development and Adolescent studies
BIOSIS Previews, 1969-present	CINAHL
Current Contents Connect, 1998-present	eBook Collection
Data Citation Index, 1900-present	Education abstracts (H.W.Wilson)
Derwent Innovations Index, 1966-present	Educational Administration Abstracts
Essential Science Indicators	ERIC
Journal Citation Reports	MEDLINE
KCI-Korean Journal Database, 1980-present	MLA Directory of Periodicals
MEDLINE, 1950-present	MLA International Bibliography
Russian Science Citation Index, 2005-present	OpenDissertations
SciELO Citation Index, 2002-present	
Zoological Record, 1864-present	

Clinical reasoning is an overarching term that is used by many authors to encompass both diagnostic and therapeutic reasoning depending upon the focus of the work in question. Critical thinking and clinical judgement are often used as terms within the literature, rather than being the focus of it. Additional appropriate papers were found through the secondary citation and grey literature review (information produced outside of traditional academic publishing and distribution channels e.g. reports, newsletters, government documents, speeches).

This initial search yielded 225 papers for review, after removal of duplicates and non-English language papers. A process of abstract review was undertaken on these papers for immediate relevance to the topic, specifically understanding the decision-making process in the clinical context. This yielded 55 papers for detailed review to provide a full view of the decision-making process.

Subsequently a search using google scholar, the grey literature, and citation review of existing references around areas drawn out of the paper review, yielded an additional 29 papers for further evaluation and incorporation into this literature review.

After initial construction of the concept framework set out in this review, a further 35 papers linked to learning concepts in clinical decision-making were re-evaluated, and 15 of these were reviewed to add breadth and depth to considerations about clinical reasoning as related to education.

Finally a further citation review (evaluation of papers cited by those already for inclusion in this work) focussed on assessment of clinical reasoning, as related to education, was undertaken yielding 29 additional papers for review to focus in on this particular area.

Papers were evaluated through the use of the SPIDER tool for paper evaluation which has been suggested to be superior to PICO when considering qualitative or mixed methods studies and thus aligned better with much of the literature retrieved in the searches (Cooke *et al.*, 2012). The search strategy is covered in Appendix 1 whilst an example of the review process for these papers can be found in Appendix 2.

Much of the medical literature has focussed upon dual process theory as the primary process by which clinicians make decisions. Elstein was a key author of this model which postulates that decisions are made through a combination of non-analytical (system 1) and analytical (system 2) processes, with the non-analytical tending to take precedence in the initial stage of the process (Elstein *et al.*, 1990). Croskerry and other authors, such as Kahneman, developed thinking on dual process theory as illustrated in figure 1 (Croskerry, 2009b; Kahneman, 2011).

2.5 Concepts of Clinical Decision-making

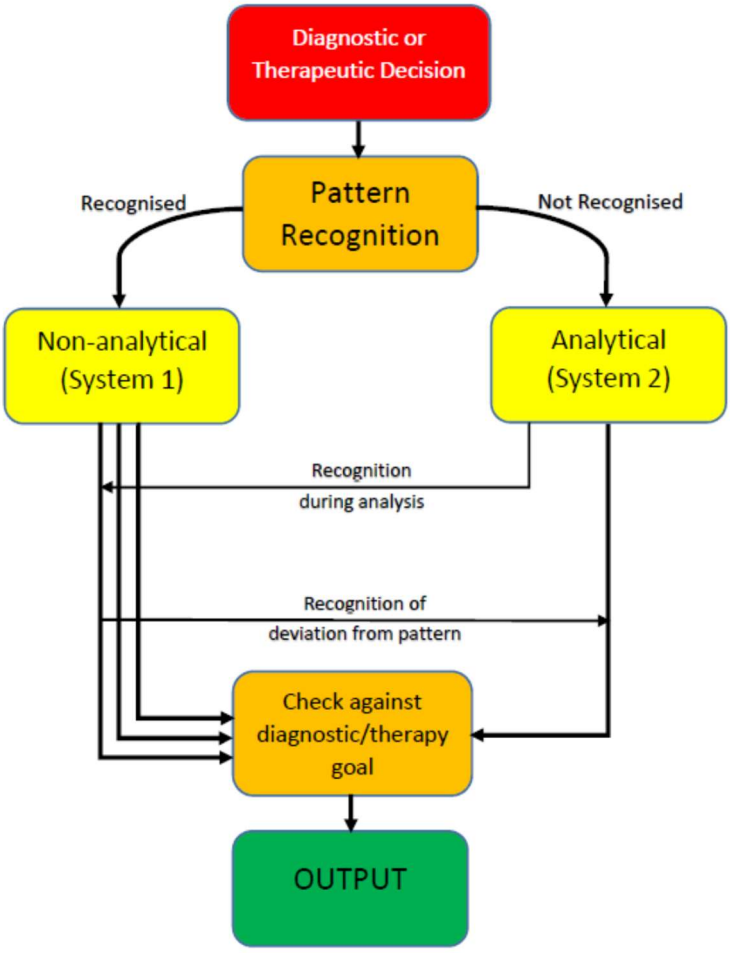


Figure 1: A dual-process model for diagnostic reasoning. After work from Croskerry P. A universal model for diagnostic reasoning (Croskerry, 2009b).

2.5.1 Concepts of Diagnostic Expertise

Dual process theory fits with concepts of diagnostic expertise, with the literature focussing on how expert decision-making tends to utilise non-analytical processes, particularly illness scripts, to reach diagnostic decisions. Evidence suggests that junior clinicians utilise biomedical knowledge to understand the presentation of the patient, however as they become more expert this knowledge becomes less relevant as their pattern recognition, or personal case experience, develops. Schmidt et al developed this further, describing a stage-by-stage process that clinicians undertake in their expertise development and the production of internal illness scripts (Schmidt *et al.*, 1990). The suggestion of a natural shift from analytical to non-analytical reasoning, with increasing expertise, has its challengers. A

study of medical students found that most used pattern recognition, or system inductive reasoning (inducing an answer from observation of specific features), rather than a hypothetico-deductive approach (formulating and testing hypotheses to acceptance or rejection) contrary to the traditional viewpoint of novice clinicians (Lopes *et al.*, 2018). The authors noted that premature closure (a cognitive bias in which alternatives are not considered once an initial decision is formed) was a higher risk in this group compared to expert clinicians, and the constructs used were based on factual data rather than experience. The same phenomenon has been observed in studies of psychology students and, challenging the simplistic binary split of cognitive process illustrated above, the evidence suggests that a single decision-making approach is worse than a suitable combined approach (Ark *et al.*, 2006). Intuitively, it certainly feels that to be deliberately making a choice to take one of two paths in clinical care, noted previously to adopt a naturally "grey" area, feels an uncomfortable concept. It is worth noting that even in his seminal work "Thinking Fast and Slow" Daniel Kahneman (2011) notes that the concept of two distinct and separate systems of analytical and non-analytical decision-making is metaphorical with no psychological reality. In later work Norman (Norman *et al.*, 2013) notes that the distinction between the two is of no more than academic interest.

2.5.2 Early Hypothesis Generation

The concept of clinical information leading to a binary position of a clinical diagnosis being "recognised" or "not recognised" is challenged by research evidence which supports the principle of early hypothesis generation in the diagnostic process. Work in the Emergency Department examined the hypothesis generation process throughout the clinical consultation. It was found that the final diagnosis was already considered at the earliest point of the clinical encounter, even prior to seeing the patient, and the diagnosis at discharge from the department was most commonly the initial one (Pelaccia *et al.*, 2014). Another study, undertaken within a primary care walk-in centre, reviewing the decision-making process through the clinical consultation, demonstrated a similar picture with examination and tests having little effect on the final diagnosis (Gruppen *et al.*, 1988). This supports work in the 1970's (Hampton *et al.*, 1975) which established the primacy of history, from which we can infer that examination and tests are most commonly part of hypothesis confirmation rather than being hypothesis generating in themselves. It is

important to recognise that the early hypothesis generation in these studies represents a non-analytical process (Custers, 2013). A mixed-methods study demonstrated that effective early hypothesis generation is linked to clinician experience with specific patient stories in differing contexts, the concept of illness scripts (Custers *et al.*, 1998). Reflecting back to the introduction, regarding the challenge of formulating coherence across decision-making concepts, the idea of hypothesis generation being a non-analytical process whilst hypothetico-deductive reasoning is an analytical process exemplifies the challenges for the learner in understanding the terminology and concepts involved.

2.5.3 The cognitive continuum and cyclical models of reasoning

Researchers agree that in most diagnostic decision-making there is an initial rapid non-analytical process that occurs with early hypothesis formation followed by further processing linked to the context in which the decision occurs. In addition the practitioner utilises heuristics (mental shortcuts to aid fast decision-making) and biases (a defined pattern of deviation from the norm or rationality in judgment) within this context to achieve a decision outcome, as described in the Emergency Department and Walk in Centre studies above. Cognitive continuum theory places analytical and non-analytical reasoning at the poles of a continuum (in their purest forms) arguing that in reality the two must act synergistically, enabling the clinician to apply each to the other in the process of decision-making within any set clinical context (Custers, 2013). Others have conceptualised the decision-making process as cyclical with an initial non-analytical process supported by an analytical "sense check" with cycling around this process until such a time as a suitable end point is reached (Marcum, 2012). This model is supported by a study involving interviews with General Practitioners which found that, whilst thinking was aligned to the dual process model, it was best interpreted as a cyclical dual process, as illustrated in figure 2, adding a reflective phase as well (Balla *et al.*, 2009).

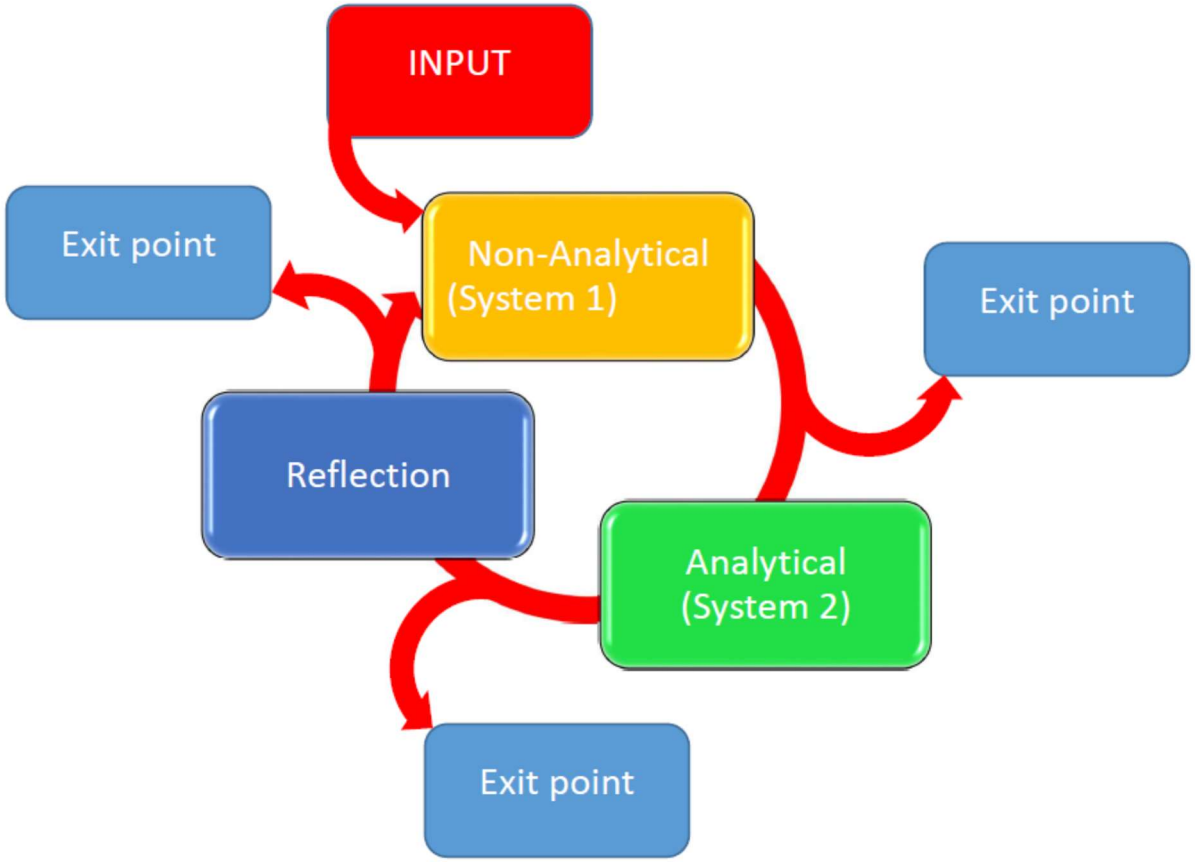


Figure 2: A conception of dual process theory utilising a dynamic cyclical model

This model is also supported by the application of the theories of Lonergan to diagnosis, commenting that diagnosis does not automatically occur as a deductive process, but rather that transitory concepts are explored and considered through a cyclical process and reflection which may yield further hypothesis and thus a cycling of the process (Malterud *et al.*, 2019). Cognitive continuum and cyclical models contrast to the dual process model in figure 1, with its implication that there is an immediate initial decision to take an analytical or non-analytical approach.

An illustrative example of the cyclical approach would be a patient attending to see a clinician with a sore throat. There is an immediate system one hypothesis formation such as tonsillitis based upon pattern recognition. This is then tested in a slower system two evaluation, through questioning to check and challenge the hypothesis, before an appropriate reflection to ensure dissonant features are not ignored. If it is felt that the

dissonant features do not suggest an alternative diagnosis then the process is exited to a decision otherwise the cycle begins again based upon the now acquired information. Alternatively the amount of initial information provided leads the clinician to join the cognitive continuum at an appropriate point between the inductive and deductive ends. As more information is gathered the decision-making flexes up and down the continuum until a diagnosis is formulated as no new data is forthcoming.

2.6 Alternative theories and concepts of clinical decision-making

As stated previously much of the medical literature, with its focus on diagnosis, has embraced dual process theory and its primary poles of inductive or deductive decision-making (Croskerry, 2009b). Other types of decision-making are described, however these other approaches may all be considered to sit along a spectrum between analytical and non-analytical. A good example of an alternative descriptor would be abduction (Ward and Haig, 1997), as well as the concepts of Bayesian logic, best represented through evidence based medicine (Braude, 2009), and the concept of categorical reasoning (Mattingly, 1991). We must also consider the place of heuristics, biases (Hamm, 2004) and bounded rationality (Simon, 1955) as these impact upon the closure point of the reasoning process. These alternative descriptors, along with these concepts around process closure are considered in this section.

2.6.1 Decision-making terms and concepts

Inductive and deductive processing sit at the opposite ends of the continuum between non-analytical and analytical decision-making. In induction, the premise presented provides some evidence for the validity of the conclusion, but not all (e.g. many patients with bowel cancer have rectal bleeding, this patient has rectal bleeding, therefore this patient has bowel cancer). Deduction is based on the assumption of two factually correct statements allowing a valid conclusion (e.g. all patients with bowel cancer have rectal bleeding, this patient has rectal bleeding, therefore this patient has bowel cancer) (Upshur, 1997). In most cases medical and psychological decision-making is considered to be abductive in nature, with the determination of decision outcome based on a set of evidence (which may be evolving) and seeking the most plausible explanation. This may be more inductive or more deductive, sitting somewhere between the two. There is an argument that medical and

psychological decision-making differs due to the understanding of the basis of any individual pattern of physical versus psychological pathology, however both can be considered to be addressed by the abductive approach (Vertue and Haig, 2008). Abduction contains elements of both deduction and induction, taking us from descriptive data patterns to recognition of specific clinical phenomena, and therefore can unite the medical and psychological approaches. It also links to the theory of explanatory breadth allowing us to select our likely diagnostic view based upon whether our produced theory (or hypothesis) explains the greatest range of facts or phenomena (Ward *et al.*, 2016). As stated previously, and covered in more depth below, these decisions are always made within a specific context (in its widest descriptive form) thus driving different cognitive processes along our non-analytical/analytical spectrum. Some authors have discussed abduction as a challenge to those who argue for the primacy of evidence based medicine which has its roots in Bayesian logic or probabilistic reasoning (Upshur, 1997). This concept utilises the probability of any particular diagnosis or treatment outcome as the primary tool in decision-making. A qualitative study with physiotherapists noted that evidence-based medicine has a probabilistic basis but is applied to an individual (Ahlsen *et al.*, 2018). They note that the decision-making process must integrate a number of sources of information and that research has demonstrated that this process is narrative based (Hunter, 1996; Greenhalgh, 1999). It can be concluded that probability has a part to play in the decision-making process, but only in so far as it is applied to the individual.

2.6.2 Bounded Rationality and Satisficing

Bounded rationality and satisficing must be included in any decision-making concept framework as these introduce the role of heuristics and biases and a consideration of the impact of context on the process. Bounded rationality postulates that as organisms we have a finite cognitive capacity and are required to make a decision based upon limited information within a finite time and therefore choose an acceptable answer within these constraints (satisficing) (Simon, 1955). The concept of the limits of human cognition is well understood and was reviewed by Young in 2014 with three key types of cognition identified -intrinsic (essential to task), extraneous (not essential to task) and germane (imposed by the learner's deliberate use of cognitive strategies). If cognitive load exceeds working memory capacity, performance and learning is impaired (Young *et al.*, 2014). This has clear

implications for the decision-making process as both intrinsic cognition (knowledge) and extraneous cognition (context) are key to effective thinking. In addition, germane cognition links to heuristics and the use of strategies to improve cognition and influence our own limited intellectual capacity through metacognition. Gigerenzer and colleagues used simulated algorithm testing to test a "take the best" approach to decision-making. Their results supported the satisficing concept statistically, whilst recognising the challenge to apply to humans due to the impact of social values and norms (Gigerenzer and Goldstein, 1996) which may influence us away from the "best" option where it conflicts with these.

2.6.3 Heuristics and Biases

The concept of bounded rationality is also supported through the use of heuristics as process management aids and, alongside them, the risks posed by cognitive biases. Early work on biases focussed on some common areas such as representativeness bias (when the similarity of diagnoses confuses reasoning regarding the probability of an outcome), availability bias (thinking of things that come immediately to mind) and anchoring bias (reliance on the first piece of information we are given), emphasising that better understanding these could improve the decision-making process (Tversky *et al.*, 1974). More contemporary authors have observed that, whilst widely criticised, the study of heuristics has not been embraced by researchers seeking to better understand decision-making (Hamm, 2004). Whilst biases arising due to the use of heuristics, as part of a non-analytic process, are frequently blamed for diagnostic error, untangling knowledge gaps and thinking errors is challenging, especially in retrospective review which is highly prone to hindsight bias (Norman, 2009). In a 2017 paper Norman et al noted that much of the work suggesting the link between heuristics and biases comes from studies in undergraduate psychology students answering expertise-free questions (Norman *et al.*, 2017). They noted the conflict between the concept that errors arise primarily within non-analytical thinking processes whilst other research suggests that it is expert clinicians who primarily utilise non-analytical hypothesis generation. It is notable that the study of heuristics is inherently challenging as it is almost impossible to establish if individuals do as they say, or say one thing then do another, and which of these is the heuristic led action (Steginga and Occhipinti, 2004). Deliberate slowing of the thought process has been postulated as a possible way to reduce diagnostic error but a 2015 study showed little benefit to this approach and it was

concluded that interventions to improve diagnostic accuracy should focus on formal and experiential knowledge (Monteiro *et al.*, 2015). Similarly a 2018 review noted a lack of long term studies showing benefit from any particular intervention and, reflecting back to the confusions of nomenclature in this topic, suggested biases and heuristics were in fact the same thing in contrast to the view of other authors (O'sullivan and Schofield, 2018). It is certainly possible to suggest that biases could be redefined as unhealthy heuristics, or indeed heuristics as positive biases, as both involve assumptions and mental shortcuts. Further work would be needed to establish if heuristics can be thought of in the same way as biases, as they would be considered as being conscious rather than unconscious, if the nomenclature is to be reconsidered.

2.6.4 Decision-making within therapeutic decision-making

Therapeutic decision-making embraces the concepts already discussed whilst introducing the concept of instrumental decision-making. In a clinical context this concept is that having achieved a diagnosis we now know where we wish to get to and thus the process is designed to get there in the most efficient manner within the context of the clinical encounter (Mattingly, 1991).

When considering decision-making within therapeutic decision-making it must be recognised that much of the research in this area has been undertaken within the therapeutic professions rather than the medical profession (who tend to focus on diagnostic reasoning). A high quality qualitative study of osteopaths found that a primarily patient centred and collaborative approach yielded the best therapeutic decision-making (Thomson *et al.*, 2014) which is in line with the work of Resnik and colleagues in physical therapists (Resnik and Jensen, 2003). In this respect the relationship between the clinician and patient acts as an enabler to the decision-making process and it is important to recognise that each interaction is unique even though the theory of treatment decision-making is usually presented as more general, an often-raised criticism of evidence based medicine and its focus on probability and Bayesian reasoning (Braude, 2009).

Linking from here into the concept of decision-making within specific context is a piece written regarding school psychologists. This notes how psychologists must consider the multiple reasons for psychological distress in a young person as well as determine why they

have a maladapted response to their personal situation thus setting the decision required into a very specific context (Andrews and Syeda, 2017).

2.6.5 Context and its impact on the reasoning process

The process of decision-making cannot be removed from the context in which it occurs, however studying the impact of contextual factors is difficult as the retrospective assessment of decisions is always limited by our inability to recreate an identical context to that in which it was made (Croskerry, 2009a). Croskerry notes that case studies seldom consider the context in which the clinical decision is made instead focussing on the case and not the "background noise" in which it occurs (e.g. location, affective state of participants), yet any decision is likely to be influenced by this external environment. An example of this is considered by Skellern when considering decisions within child protection cases in which there is a significant emotive context which can lead to biases due to the failure to manage one's own thought processes (metacognition) risking an erroneous outcome with potentially significant consequences (Skellern, 2020). Evidence shows that expertise is highly context specific with skill decay a natural consequence of a change in task or practice environment (Ericsson, 2007). In this respect expert use of heuristics must also be linked to their use in a specific decision context and thus, logically, heuristics must become increasingly prone to bias when utilised in a context away from the natural comfort zone of the clinician such as a hospital clinician working in primary care or vice versa (Weaver *et al.*, 2012).

Context is not only the physical environment in which a decision is made but also part of the natural condition of both the clinician and patient involved. A study of physical therapists found that expert therapists had a broad knowledge base fuelled by a drive for lifelong learning and patient centeredness in their practice (Resnik and Jensen, 2003). Another study of occupational therapists looking at therapeutic decision-making specifically noted the "personal style" of the individual clinician, and its influence on decision outcomes, which links to values and attitudes in the individual practitioner (Gibson *et al.*, 2000). Patient values and attitudes also contribute to these contextual factors both directly, through adherence or not to treatment, but also to the impact they have on the clinician which can then influence the decision-making process. Mjaaland et al reviewed secondary care consultations and found that when patients expressed negative emotions physicians would

move away from emotional communication thus the patient behaviour was likely to lead to direct influence on the reasoning process (Mjaaland *et al.*, 2011). Both the clinician and patient contextual factors can be linked back to a behaviourist psychological perspective which considers that both parties' behaviours affect one another's cognitive processes subsequently influencing outcomes in terms of both thoughts and behaviours (Durning *et al.*, 2013). In work using videotaped consultations with varying changes in contextual factors (patient, encounter and/or physician) researchers were able to demonstrate the impact of contextual factors on expert physician performance with more impact on diagnostic than therapeutic decision-making, possibly due to the impact of instrumental thinking on the therapeutic decision-making process (Durning *et al.*, 2012b).

Context can be brought together in situativity theory, well summarised by Durning and Artino in AMEE guide no 52 (Durning and Artino, 2011). This broadly states that all decisions and learning are situated in experience where experience is defined by the concepts of context illustrated above. If we accept this, then the impact of context must be embedded within any concept framework of decision-making as the cognitive process is fundamentally linked to the external factors in all their forms. This clearly has significant implications for medical education.

2.7 Proposing a concept framework for decision-making

Bringing together these differing concepts and components of decision-making, as expressed in the literature explored above, leads to a concept framework as set out in figure 3, a development from those previously proposed.

In this concept framework an initial input of information leads to us dropping onto a cognitive continuum at a point determined by the context in which the decision is being made and the type of decision to be made. A diagnostic decision might start toward the analytical end whilst a therapeutic decision may start more toward the non-analytical end, the exact position being determined by the context (physical, affective and urgency) in which the decision is required to be made. The decision-making will flex back and forth along the continuum as patterns are recognised or challenged with reflection on the case, until such a time as an appropriate output decision is made, which will be made within a position of bounded rationality.

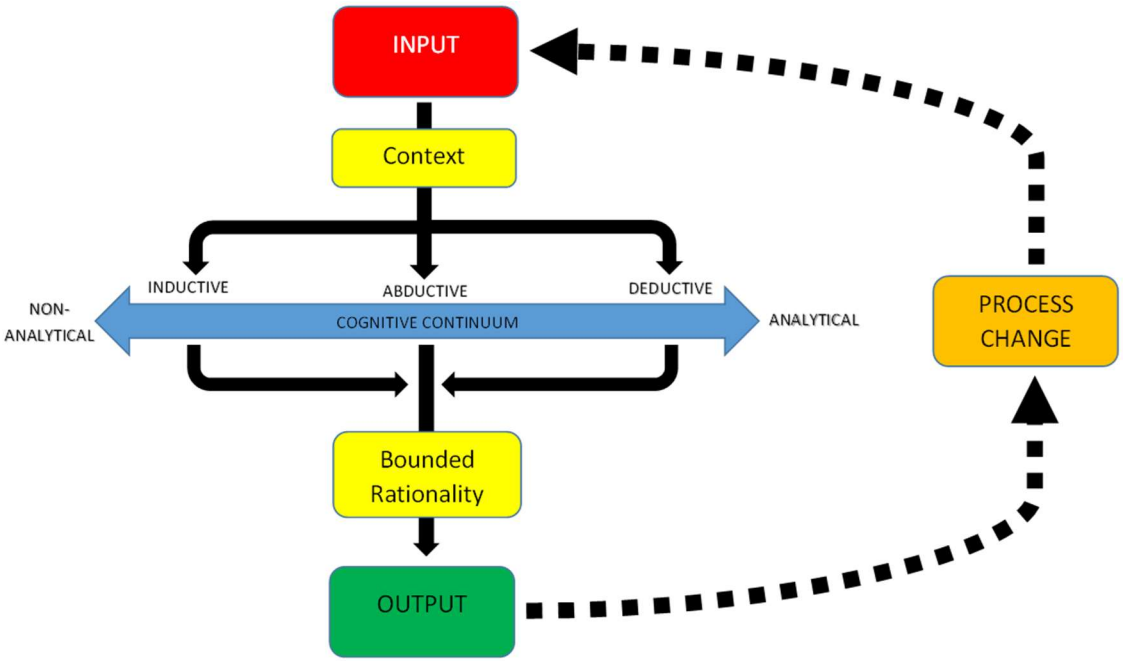


Figure 3: A concept framework of decision-making based on the literature evidence.

We then enter a cyclical process whereby a change in the patient's condition, the decision to be undertaken, or any other shift in the patient or context which requires a new decision to be made, takes us to the beginning of the process once more. In each case the context factors will be slightly different as will the constraints forming the bounded rationality of the decision rendering each individual decision-making process unique.

2.7.1 Illustrative case studies

The concept framework can best be illustrated through some example fictional case studies.

Case One

A patient is brought into the Emergency Department with crushing central chest pain. The context here is of an emergency and a rapid need to intervene meaning that the attending clinician meets the cognitive continuum at the non-analytical end formulating an early diagnosis of a myocardial infarction (MI). Bounded rationality means that they give aspirin immediately, determining that this is a good enough decision before awaiting any other tests as the benefit outweighs the risk. The tests coming in now change the situation and we move back to the top of the cycle again. The tests confirm an MI and now a decision must be made as to treatment thus there is a shift to therapeutic decision-making. Instrumental

thinking points us to primary percutaneous coronary intervention (PPCI) and our decision is predominantly non-analytical as there is a clear guideline for this. Some decision-making will be required to ensure that the patient has no contraindications and is in time for the treatment but if we can satisfice the outcome of benefit outweighing risk, we proceed to a treatment output.

Case 2

A patient attends General Practice with symptoms of fatigue over a period of several months. Here the context is a non-acute problem with many possible causes as well as many potential contextual factors pertaining to the patient (e.g. mental health issues, cultural issues, social issues) and the clinician (e.g. views of some specific tiredness syndromes such as Myalgic Encephalomyelitis). In this case there is likely to be some early hypothesis generation however the continuum will likely be entered more centrally or toward the analytical end dependent upon the amount of information provided by the patient. During the consultation additional information gathered will inform the decision-making through an abductive process until bounded rationality leaves several potential diagnoses. The output here will likely be a differential diagnosis with hypothesis testing through a selection of investigations and a plan to re-evaluate with the results. We now go round again and enter a second round of diagnostic decision-making informed through the results. These results will determine where this next round of reasoning starts on the continuum depending on their specificity. This cycle may need to be repeated several times until the clinician satisfices a likely diagnosis and can begin therapeutic reasoning. In this case the therapeutic decision may be more challenging, especially if a more psychological cause is felt to be the issue, and thus patient context may be very strong around opinions of different treatments and interventions. It could be argued from this that in therapeutic decision-making the patients will also undergo a process of bounded rationality and satisficing in accepting any specific course of action and this is where the shared decision-making process embeds itself in the process.

These two simple case studies illustrate how this model can be applied to clinical care in very different scenarios and be applied to both therapeutic and diagnostic decision-making. As I move on to consider the implications for medical education and assessment it is also important to consider this model as having the reasoning process centred on the patient. In

doing so it allows a description of the process to be more easily conceptualised within any specific context. In order to consider these impacts it is important to have a clear operational definition of the term clinical reasoning.

2.8 An Operational Definition of Clinical Reasoning

The defined concept framework gives an opportunity to provide an operational definition of the term "clinical reasoning" to be utilised through this work. As stated early in the chapter there is a lack of coherence in the use of the term and thus in any work on the topic it is important to be clear on the definition in order to aid the readers understanding and permit effective understanding of the work and its implications more widely.
The concept framework development, through this narrative review, has deliberately utilised work from across different clinical roles and specialities in order to facilitate a wider understanding of the subject. It has been constructed recognising that decision-making in the clinical environment involves decisions of both diagnosis and management and that whilst the latter are aided by a recognition of the intended endpoint, both represent a distinct process.

For this work therefore the operational definition of Clinical Reasoning is all decisions taken by clinicians in reaching diagnostic, and subsequently therapeutic, decisions. Whilst these are two processes, they can both be explained effectively by the constructed concept framework defined above, as illustrated in the case studies. From this point forward in the work the use of the term clinical reasoning will refer to this definition however recognising that some literature uses this term in a different way, or one of the other terms previously mentioned, the term decision-making will be used where the above operational definition would be considered inappropriate or not explicitly correct.

2.9 Implications for Clinical Education

With effective decision-making recognised as key to improved diagnostic and therapeutic outcomes it is important that it is addressed within clinical education. Evaluation of concepts of decision-making across four professions (medicine, nursing, pharmacy and social work) yielded consensus that decision-making should be embedded in education however how this should be done could not be agreed upon (Kahlke and Eva, 2018). Educationally this is important as it makes the teaching of any one approach, to the

detriment of others, both unrealistic and a failure of our duty toward students. Norman commented as such in 2009 noting that it was difficult to be prescriptive about which educational approaches would be effective to develop decision-making skills (Norman, 2009).

In a 2020 paper Richards et al observed that medical schools have traditionally focussed on the "knowing" rather than the "thinking" of clinical care whilst taking the position that the skills to make effective decisions can be taught through processes typically delivered whilst in the clinical environment such as the one-minute preceptor, a micro skills model probing the decision-making process of the trainee in reaching a particular conclusion (Richards *et al.*, 2020). The authors acknowledge that such instructional approaches still have to be done well in order to prevent reliance on illness script (inductive) approaches as the overall time spent on the clinical encounter will outweigh the instructional time and thus the core cases impact on the student might outweigh the impact of the instruction. Previous work with medical students found that they made decisions in the same way as doctors and so we must ensure that we engage with the pattern recognition approach and its initial triggering of hypothesises rather than focus on hypothetico-deductive approach as the sole way to teaching decision-making (Neufeld *et al.*, 1981). A more recent literature review concluded that an early introduction to illness scripts is likely to benefit trainees with teachers defining scripts but the scripts themselves being constructed by the learner enabling them to create their own pattern recognition system within defined parameters (Charlin *et al.*, 2000).

2.9.1 Teaching clinical reasoning

In their 1981 paper Neufeld and Norman concluded that we cannot view decision-making as a general skill to be taught and evaluated as it is multifactorial in nature (Neufeld *et al.*, 1981). As educators we must be prepared to acknowledge the challenges in providing a "one size fits all" approach to teaching a subject which is the exact opposite of this. As Aristotle stated in his Nicomachean Ethics, "We should not expect more exactness from a subject than the nature of the subject allows" (Aristotle, 1926).

It is important that we link the clinical knowledge and practical delivery of clinical care through decision-making, especially if we are to produce adaptive practitioners who can be responsive to the clinical case that presents regardless of aetiology. In an editorial

Sockalingham et al. note that experts have efficiency dimensions (knowledge to solve routine problems) and innovative dimensions (creating new solutions to difficult problems) yet clinical education is focussed on the former potentially stifling the creation of such adaptability (Sockalingam *et al.*, 2016). Adaptive practitioners define excellence thus we must ensure that as well as the provision of knowledge we provide the skills to utilise it, including how it changes during its application (Mylopoulos *et al.*, 2018). Work by Geisler and colleagues considered the alignment between views in diagnostic and therapeutic decision-making. They emphasise the need for educators to connect the knowledge acquisition to its application in therapeutic decisions as although the addition of instrumental thinking at this point aids us (as do guidelines) they are still being applied within the context of an individual patient (Geisler and Lazenby, 2009) as previously noted in work from physiotherapists.

It should be considered that the use of teaching on specific concepts of decision-making such as probabilistic (x more likely than y), causative (mechanistic description of dysfunction) and deterministic (if x then conclude y) are difficult to operationalise in decision-making within clinical interactions. It is due to this that the hypothetico-deductive approach and script theory have been more commonly taught directly as concepts (Stempsey, 2009). Balla and colleagues recommend that education should focus on building mental models through experience of multiple cases (Balla *et al.*, 2009) and there is a body of research that focusses on concepts linked to case discussion and classroom based approaches to decision-making. These approaches include using similar cases and drawing out commonalities and differences (Speicher *et al.*, 2012) as well as systematic deconstruction of cases post-hoc to re-evaluate the information and re-explore it (Delany and Watkin, 2009).

We must ensure that any curriculum and assessment development in clinical reasoning is context specific, with a focus on high numbers of clinical encounters, as skilled intuition (and thus primary non-analytical hypothesis generation) can only occur with sufficient regular exposure within the environment to which it pertains (Kahneman and Klein, 2009). On this basis it is also important for us to instil in trainees that this will be a lifelong undertaking, and caution must be applied when they move specialities in their undergraduate and early postgraduate education. It must also be recognised that no two clinical students can have

the same educational experience due to the varied patient exposure that each will receive, even in the same clinical speciality (another context factor), and thus we need to ensure our approaches to learning acknowledge this (Eva, 2005). Pelaccia, writing specifically on dual process theory and medical education, noted the need to intertwine decision-making from the start of the education process with the encouragement of immediate feedback through direct observation and early questioning in the clinical environment through effective supervision and support (Pelaccia *et al.*, 2011). Importantly the traditional approach of sending a student to see a patient then report back with an exhaustive clerking will not yield effective improvement in a student's decision-making ability and we should be moving away from this approach (Norman *et al.*, 1999). Whilst this is a 1999 paper, and things have developed since then, the busy nature of the clinical environment means this can still be a common occurrence rather than a process focussed on "why" questions pertaining to the way the information gathered guides us to a diagnostic or therapeutic decision.

2.9.2 Core skills and approaches to teaching clinical reasoning

The concept of producing a diagnosis as a core skill may also need to be explored in more detail. Whitbeck reflects that diagnosis itself has little value, rather the focus should be on the management of the disease entity not the ability to name it and thus reframing the exercise of diagnosis to the process, rather than the outcome, may add value to clinical education (Whitbeck, 1981). A study conducting semi-structured interviews with General Practitioners noted the need to consider the exclusion of red flags alongside the pursuit of diagnosis to ensure the development of safe practitioners (Balla *et al.*, 2009). It is important to ensure that we do not simply focus on diagnosis and its empirical accuracy but also the credibility of the clinical reasoning process through its coherence in relation to the case to be evaluated. Authors have suggested the use of fully blinded cases (diagnosis unknown to all participants) to facilitate this as it can prevent facilitators consciously, or unconsciously, pushing a particular line of thinking on students (Custers, 2019).

Communication skills are another core area which must be considered as a building block of decision-making education, but it cannot follow that a single approach can be taught which works for all contexts. With the consultation conceptualised as a dynamic process, communication skills are vital to facilitate the rule in, or out, process of diagnostic decision-making (Malterud *et al.*, 2019) as well as the shared decision-making approach needed for

effective diagnosis. Cox et al considered the medical consultation through language and social interaction theory and considered it to be an inherently creative event and comment upon the contextual implications highlighting the risk of misalignment if there is a failure to adapt to the patient and their nature (Cox and Li, 2020). This is linked to the concept of patient ideas, concerns and expectations and the risk of poor consultation outcomes when these are not addressed effectively by the clinician who instead focusses on their own agenda. Kassirer notes that cognition is only one element of the decision-making process with communication and information extraction also crucial (Kassirer, 2010). A nursing study supports this concept noting that language permits a full interaction between humans and thus a failure to align this between clinician and patient (both within and across individual languages and non-verbal cues) is likely to lead to decision-making deficiencies (Ceolin *et al.*, 2017). This can also link to challenges encountered when either party in the interaction does not enact or process these normally such as the neurodiverse or those with other physical or mental disabilities.

A number of authors have written of approaches to embedding the teaching of decision-making into clinical education including use of fuzzy trace theory (pattern recognition based upon two forms of recall – verbatim and gist), through use of online resources (Lloyd and Reyna, 2001), the use of thinking frames (Neistadt, 1998), and pause and reflect within simulations (Pennaforte *et al.*, 2016). Review of these papers finds that most are low powered studies or process descriptions with little hard evidence to support improvement in decision-making skills further emphasising our difficulties in developing unified approaches to delivering teaching on the topic within our curricula. It is important that we continue to build our evidence base in this area to support clinical educators as we have the responsibility to produce the next critically thinking clinicians (Barrett *et al.*, 2018).

2.10 Assessment of Reasoning

Assessment of clinical reasoning within existing, commonly used, assessment methods has been an ongoing challenge. As suggested previously, for medical education to respond effectively to produce a true holistic assessment of an individual clinicians performance (to include the assessment of clinical reasoning) a more programmatic approach to assessment must be employed with multiple data points, purposive sampling and ongoing learner

feedback (Van der Vleuten *et al.*, 2012; Schuwirth *et al.*, 2017). If this is to be realised then we must better understand the way clinical reasoning occurs across multiple assessment methodologies and how we can assess it within our marking schemes. The views of these authors is supported by thinking in postgraduate education where Emergency Medicine physicians have recognised a consensus view that multiple assessment strategies are required for decision-making skills as one alone has insufficient validity (Ilgen *et al.*, 2012) a position echoed by other authors within undergraduate education (Epstein and Hundert, 2002).

2.10.1 Written Assessments

The most widely used assessment methodologies centre on key feature question approaches which include single best answer/multiple choice questions, extended matching questions and others, with the underlying concept being that relatively few key clinical elements are required for resolution of a clinical problem be it diagnostic or therapeutic. This leads to the argument that they cannot effectively test decision-making as any information in the question is deemed relevant, and therefore important, in the "correct" outcome (Hrynchak *et al.*, 2014). The suggestion that we might choose to move away from key feature question approaches however needs to be set against the recognition that these assessments give educators the ability to sample broadly across a curriculum in a relatively simple way (Daniel *et al.*, 2019). The view that these assessments do not test decision-making is challenged by earlier work which looked at modified essay/short answer (MEQ/SAQ) and multiple choice (MCQ) questions and found that many MEQs simply measured recall and facts and that with good exam writing support MCQs were actually better at testing higher level cognitive reasoning skills (Palmer and Devitt, 2007). It may therefore be that we need to provide more explicit and objective guidance for question writers, with well-developed quality assurance methodologies, to get the best assessment of clinical reasoning within these modalities.

An alternative option, designed specifically for written testing of decision-making, is the script concordance test which utilises an approach whereby the introduction of vignettes of information is followed by a student indicating if it makes a diagnosis more or less likely and is designed to work in testing decision-making in situations of uncertainty. It has some methodological drawbacks in that 15 to 20 raters are needed to create the benchmark

scores, and it has tended to be diagnostically focussed rather than therapeutic with concerns also raised that it oversimplifies knowledge (Esteves *et al.*, 2013). Of greater concern is work demonstrating that simple exam technique can significantly manipulate scores and that there are notable measurement errors in work promoting the methodology. Like the work on key feature questions the authors suggest alteration to the marking of these, and improvement in the question formulation, may yield positive outcomes (Lineberry *et al.*, 2013).

2.10.2 Objective Structured Clinical Examinations

Objective structured clinical examinations (OSCE) seek to assess at the "shows how" level of Miller's pyramid (Miller, 1990) and as such have the potential for greater assessment of clinical reasoning skills. Authors have sought to develop an understanding of decision-making through the use of post-encounter forms, essentially a two part OSCE station in which the second part necessitates students to write their differential diagnoses and the information gathered which either supports or refutes the diagnosis (Durning *et al.*, 2012a). This approach was explored further in a study from South Korea where the calculated decision-making scores from the post-encounter forms were linked to different known academic scores for each individual student. They found no correlation with OSCE or knowledge test score but there was between the individual grade point averages for the students and the diagnostic accuracy score. This strengthens the argument that programmatic assessment is needed to adequately evaluate decision-making skills (Park *et al.*, 2015).

The other key aspect of the OSCE is the marking process, specifically the use of checklist marking or global domain marking. A case-control study examining the difference between the two found that global rating based schemes had higher reliability and construct validity however the study is weakened by having the same assessor complete both marking schemes with the risk that completion of one leads to bias in the completion of the other (Hodges and McIlroy, 2003). Checklist marking is also questioned by Schuwirth, who comments that checklist marking penalises efficiency in the diagnostic process and is therefore at odds with the concepts of expertise and the use of non-analytical decision-making approaches (Schuwirth, 2009), whilst Thampy et al note that checklists trivialise the complexity of the clinical interaction (Thampy *et al.*, 2019).

2.10.3 Workplace Based Assessment

Consideration of the assessment of clinical reasoning must conclude with consideration of workplace-based assessment (WBA) processes. An important link between OSCE and WBA is the need for effective alignment between the rating scale and the judgement being asked. A large multicentre trial demonstrated an improvement in assessor decision-making in WBA when this alignment was in place (Crossley *et al.*, 2011). This alignment has been commented on by other authors as having the potential to improve the assessment of OSCE with global rating scores by reducing bias inherent in checklist mark schemes, especially if used with borderline regression methods. There remains a risk that the constant problem of inter-rater reliability may be amplified with ordinal scales (rather than more nominal checklists) as early categorical judgements made by assessors can influence where on scales they anchor their judgements so continued refinement of the judgements to be made is vital (Gingerich *et al.*, 2011). WBA methods support the concept of time variability for competence acquisition and that competency based medical education cannot be supported by scheduled assessment alone as different individuals will acquire skills, including decision-making skills, at different rates. Multiple formative assessments can enable us to make an effective summative judgement and, if we are to make effective judgements regarding decision-making, we should also ensure a broad sampling across time, speciality, and case to remove context as a potential bias to our assessment processes (Gruppen *et al.*, 2018).

Multiple tools have been suggested to facilitate WBA processes. Early studies and practice have established the Mini-CEX as a superior methodology to facilitate multiple encounter and context sampling of clinical practice through evaluation of multiple encounters (John *et al.*, 2003). A 2009 systematic review supported the high validity and reliability of the mini-CEX however commented that observers often lacked adequate training to use the tools and greater faculty development is required for quality assurance of these assessments (Kogan *et al.*, 2009). A 2019 concept analysis noted that nursing is more prone to try to develop thinking rubrics to facilitate decision-making evaluation (Manetti, 2019) particularly noting Lasater's clinical judgement rubric which was developed through the assessment of simulation (Lasater, 2007). Others have sought to develop assessment tools within medical education specific to decision-making. Examples include the "assessment of reasoning" tool

(similar in approach to mini-CEX) designed for use with case based discussion and workplace based assessment (Thammasitboon *et al.*, 2018) or the IDEA tool designed to evaluate decision-making as displayed in the written record through a structured review proforma (Baker *et al.*, 2015) however the assessment of these in practice is more limited. Development of "talk aloud" processes to understand the decision-making undertaken by an individual have been studied in theoretical cases although it was noted that significant training was required to implement successfully. Part of the issue raised by authors was whether a decision-making process can be considered bad if the problem is solved and whether pattern recognition is side-lined by the talk-aloud process (Pottier *et al.*, 2010). It could be argued that talk-aloud is part of the widely used one-minute preceptor model under the second described micro-skill of "probe for evidence" (Neher *et al.*, 1992). It may be that a future development for formative assessment would be to develop the training of the preceptors to enhance and rate this aspect of the interaction to create multiple micro-assessments of a student's clinical reasoning ability.

2.11 Summary and formulation of my research question

The role of those developing clinical curricula must be to ensure that there is constructive alignment between clinical reasoning within the taught curriculum and its assessment. This is best achieved through considering the scheme of assessment at a programme level rather than through its individual components (Biggs, 1996; Van Der Vleuten and Schuwirth, 2019). If we accept the principle that "assessment drives learning" then ensuring assessment processes effectively evaluate clinical reasoning can facilitate curricula change, driven by both educators and students, with a better understanding of how students undertake clinical reasoning within assessments likely to yield significant value for our pedagogical practice.

Assessment remains an area in which clinical reasoning skill continues to be difficult to evaluate. Whilst researchers have sought to develop tools to evaluate decision-making skills through assessment, there remains a lack of clarity as to the way students actually undertake this process within assessments. Putting context as a primary factor within the process of clinical reasoning, as set out in the concept framework, means we need to understand the context of being in an assessment and its impact on the clinical reasoning

process. Presently there is little understanding of how the context of assessment may alter student thinking and whether it is at odds with the information we have from previous research in the workplace. If there is a clear dissonance between the two it leaves a gap that we, as educators, need to be able to close through curriculum development and reconsideration of our assessment approaches.

This leads to the research question to be answered which is “How do students undertake the process of clinical reasoning in Single Best Answer questions and Objective Structured Clinical Examinations?”. I seek to understand this process better in order to enable clinical educators to develop a more programme-based curriculum and assessment approach that can ensure that we do not disadvantage clinicians’ development of effective reasoning whether they utilise analytical or non-analytical approaches.

Chapter 3 – Underpinning methodology

3.1 Introduction and Chapter Summary

This chapter will set out the basis of the methodology to be used within this research study. As this study utilises a mixed methods approach as its primary approach, the underpinning assumptions of this will be discussed initially considering the rationale for its use in this particular research project.

The following part of the chapter will look in more depth at the individual quantitative and qualitative components as these form the basis of the mixed methods approach and thus assumptions regarding them must be made and addressed. These two components will then be brought together in a discussion of the mixed methods methodological approach to be used through data triangulation and interpretation within this work that will produce the final results.

The chapter will close with a discussion of consensus group methods, as a nominal group approach is used to define the questions for the study overall and therefore an important contributor to overall study validity and thus must also be clearly defined and considered.

3.2 Mixed Methods Research

Mixed methods research is defined as the use of qualitative and quantitative data in the methodology of a study in order to gain deeper understanding of the topic at hand through the ability to gain both breadth and depth of knowledge in the topic (Creswell and Clark, 2017).

This study will use a mixed methods approach to address the question as to how clinical students undertake clinical reasoning within assessments. Medical education research fits within the social sciences more than typical medical model (biomedical) research and thus, in many cases, is not well suited to purely quantitative methods, despite these being generally favoured by clinicians, as they align to the evidence based practice concepts espoused by many in the profession (Evans and Benefield, 2001).

Mixed methods research is of particular value in medical education research due to the complex nature of interactions within the education encounter, not dissimilar to that in the

primary care clinical interaction, an area considered ideal for mixed method research by key subject matter authors (Creswell *et al.*, 2004). With both primary care and medical education involving social and clinical elements to undertake clinical decisions, the potential value of mixed methods to address this research topic is clear.

3.2.1 Philosophical Assumptions

In developing any research study it is important to first understand the philosophical assumptions which provide an overarching structure to the study and guide the development of the methodology. Worldview is an alternative way to consider the concept of a paradigm defined originally by Kuhn as "a set of generalisations, beliefs and values of a community of specialists" (Kuhn, 1996). Although Kuhn's work has been criticised by more contemporary authors, the concept of a worldview (as proposed by Creswell) remains important. Mixed methods research is most aligned to the pragmatist worldview, focussing on the way the methodology allows us to address the research question within the real world, utilising quantitative and qualitative approaches that work most effectively to achieve this (Creswell and Clark, 2017; Christensen, 2022). The adoption of the pragmatist worldview allows us to address a number of key philosophical questions which shape the proposed mixed methods research methodology.

Ontology is concerned with the nature of reality and relationships between different concepts of reality. The pragmatist worldview recognises that there are both singular (hypothesis accepted or rejected) and multiple (different perspectives) realities rather than focussing on one or the other. Positivism promotes the single worldview and lends itself well to quantitative research whilst constructivism promotes multiple world views and is associated with qualitative methodologies. Neither on their own can facilitate the understanding of a concept such as clinical reasoning which cannot easily operate within a position of right and wrong, nor in a position in which a final decision can be left as a broad concept.

Epistemology defines the limits of knowledge and, in research terms, the relationship between the researcher and that being researched (Creswell and Clark, 2017). Positivism advocates the researcher being remote from the research process, a passive observer, whilst the constructivist approach centres the researcher within the process guiding the

understanding and data capture. The conflict between these approaches both challenges, and advocates for, the mixed methods approach under a pragmatist worldview (Poses and Isen, 1998). Applying this worldview it can be seen that, with the focus on the methodological approach to achieve the outcome, we can utilise both quantitative and qualitative approaches effectively to obtain better quality results.

Finally this leads us to the Axiology of the research, defined as the role of the researcher's own values on all stages of the research process. The pragmatist worldview allows us to recognise the benefits of both biased, and unbiased perspectives within the study, improving our understanding of the topic at hand. The pragmatist worldview has been advocated as most in line with the mixed methods approach by key subject authors and fits well with the question being asked in this study (Tashakkori and Teddlie, 2010).

3.2.2 Quality and validity in mixed methods research

Mixed methods research validity has been subject to significant scrutiny and consideration due to its approach of utilising both qualitative and quantitative methodologies. O'Cathain describes three approaches to assess quality within mixed methods studies, concluding that a bespoke mixed methods design is crucial, notwithstanding the need to address quality within the specific qualitative and quantitative components (O'Cathain, 2010). Onwuegbuzie and Johnson take this further, preferring to use the term "legitimisation" rather than validity, seeking to provide a term to straddle both research methods. They also note that the iterative and interactive nature of mixed methods research means that the ability to be definitive about any inferences made may never be achievable (Onwuegbuzie and Johnson, 2006).

Creswell and Plano Clark recognise that as each mixed methods study has its own specific purpose and design the threats to validity vary according to the specific design being utilised. This study utilises a convergent design and it is therefore necessary to ensure appropriate validity in the individual quantitative and qualitative components of the research (Creswell and Clark, 2017). Finally it is clear that matters of validity within mixed methods is particularly focused on the data integration which forms a fundamental part of any mixed methods study and differentiates it from other methodological approaches. The

convergent design, and its approach to data integration, will be discussed further in this chapter.

It can be seen that in order to effectively understand mixed methods research we must first understand the components and theory underpinning the quantitative and qualitative methodologies used in the initial data collection and interpretation.

3.3 Quantitative Research Methodology

3.3.1 Quantitative Research theory

Quantitative research has an underpinning positivist worldview in which the researcher is detached from the observations that they are making. It has been the dominant paradigm in research over many centuries since the work of Descartes (Descartes, 1649). Epistemologically it focuses on the objective collection of data with the researcher undertaking data collection remotely without their opinions contributing to the outcome, thus seeking to minimise any subjective interpretation of the data (Tavakol and Sandars, 2014a).

Ontologically it supposes there to be only one reality, or truth, which can be observed through cause and effect, thus the need for the researcher to be a “cold” observer within the process. This links to its methodology in which a theory is developed into a hypothesis which is then tested to seek evidence of statistical probability. The positivist worldview ensures that generalisability of the results can be assured. The ontological position of a singular truth means that any similar situation must be assumed to function within this single reality with a similar measurable outcome (Denzin and Lincoln, 2011).

It is important to recognise the criticism that the positivist worldview has had, particularly in the context of its application to social and educational research. Post-positivists argue that absolute reality cannot be ascertained and thus only an estimation can be made. These research commentators, many being the drivers of the move toward mixed methods research, argue that multiple methods are needed to obtain a view of the reality being tested and that theories require continual refinement and retesting. Nevertheless those advocating the post-positivist worldview still work from theory, to observation, to outcome and with a focus on objectivity. This describes a deductive approach and thus the

quantitative viewpoint is still predominant rather than a more inductive approach aligned with qualitative research (Tavakol and Sandars, 2014a).

3.3.2 Ensuring validity in quantitative research

Within quantitative research the measure of validity is linked to the positivist worldview recognising that if there is accepted to be a single reality which the research seeks to elicit then the outcomes should be seen to be consistent and predictable between samples (Cohen, 2018). Validity is therefore underpinned by the statistical methods utilised to evaluate the data and the assumptions that are made within the method regarding the data sample itself (linear v non-linear, normally distributed v non-normally distributed). The approach to statistical evaluation is discussed in more detail below, however it is clear that ensuring appropriate sample size is a key component of the validity within the quantitative arm of the study.

The other aspect of validity within the quantitative aspect of the study will be ensuring that the questions asked of the students are appropriate to answer the question being asked within the research (Tavakol and Sandars, 2014b). It is for this reason that the Nominal Group Technique (discussed below) will be used to ensure that there is appropriate expert scrutiny of this aspect of the work to provide this validity assurance.

3.3.3 Sampling decisions in quantitative research

Sampling is the action of selecting a suitable research group from a wider population. The aim of quantitative studies is to gain a broad understanding of the phenomenon at hand. This requires relatively large sample sizes, especially when dealing with non-normally distributed populations or data sets as in this study (Cohen, 2018). Large data sets can reduce the risk of sampling error where the result from the measure group differs from that of the entire group under evaluation, for example, the observed group of medical students versus all medical students. It is notable that sampling error is considered to be difficult to account for and not within the ability of the researcher to influence (Tavakol and Sandars, 2014b). Larger samples can also facilitate reduction in the risk of sampling bias which is where the sample studied is unrepresentative of the wider target group against any given characteristic (sex, age etc). In the context of this study a larger sample size in the

quantitative element of the study will ensure higher representativeness within the sample and reduce the risk of bias.

In summary, decisions regarding the exact number of participants required requires determination through consideration of the number of variables being evaluated and the need to reduce the risk of sampling error and bias. Discussion of sample size and its calculation is covered in chapter 5.

3.3.4 Limitations of Quantitative Research

It is important to recognise the inherent challenge in undertaking a traditional quantitative approach to the evaluation of a reality where the objective truth sought is observed through the eyes of the individuals under evaluation such as in this project (Cohen, 2018). Whilst the use of a quantitative approach, with the researcher gaining independent and objective data over which they have little influence, is of significant value the phenomenon under study does not fit simply with the positivist approach of observing the influence of an independent variable on a dependent one with a singular reality. This represents the primary limitation of quantitative research methods. The reduction to the single reality, and the consequent need for statistical analysis, can reduce the validity of the findings if the question being asked does not fit comfortably within this paradigm. This supports the post-positivist view point of a more estimated research conclusion and the value of mixed methods for approaching projects such as this where it needs to be acknowledged that individuals will experience the phenomenon from their own specific reality (Atieno, 2009).

3.3.5 Undertaking Data Analysis in Quantitative research

Data analysis in quantitative research relies on statistical analysis with methodology based upon the nature of the data under evaluation. In this project it is proposed to use Likert data items within the quantitative arm of the study in order to gain a contemporaneous view from students (Likert, 1932).

There are both advantages and disadvantages to the use of Likert data items. Advantages include their simplicity of construction, the likelihood of a reliable scale, and the ease of use by study participants. Conversely, they can tend toward central tendency bias, with participants avoiding extreme responses, and both acquiescence and social desirability bias,

where individuals will seek to answer either in the way they think the researcher wishes or a way they feel is more in line with social acceptability, despite it not being their true view.

The resulting data will be ordinal and discrete and thus would not be considered to meet the criteria required to permit analysis using parametric data analysis methods due to a lack of normal distribution. Instead it is usually considered appropriate to use tabulation, frequency methods or non-parametric statistical analysis (Allen and Seaman, 2007). It should be noted however that whilst the data is not normally distributed and therefore non-parametric statistical tests would be considered more appropriate there is some debate in the literature as to whether Likert data can be analysed with parametric statistical methods. Likert scales are termed ordinal data and therefore the assumptions of parametric tests are not met, however when the scale is from 1 – 10 (strongly disagree to strongly agree, as in this case) then some argue that it can treated as interval data (equal distance between data points) and therefore parametric test assumptions can be met. Also assuming that a large sample size of greater than n >= 200 would be of a normal distribution then it can be argued that parametric tests can be used in this test study (de Winter and Dodou, 2010). Following the construction of the data collection tool it is proposed that the first data collection will undergo pilot analysis in order to establish the most appropriate data analysis approach prior to its application to all data. In the use of these scales each can be evaluated using the relevant statistical testing against the hypothesis inherent in the question prior to conclusions being drawn. SPSS will be utilised for data analysis to ensure data evaluation is robust and several approaches will be used to ensure that the conclusions can be considered reliable (de Winter and Dodou, 2010).

3.4 Qualitative Research methodology

Qualitative research methods are grounded in social reality and allow us to explore questions for which the formulation of a hypothesis is difficult. The essential features of such research focuses on choosing an appropriate method, recognising and analysing the perspectives of the participants, and has the researcher as an integral part of the process concerned with knowledge production (Flick, 2018).

3.4.1 Qualitative Research Theories

Qualitative research represents an umbrella term for a number of different research approaches; however most are focussed on a constructivist (also known as naturalist) worldview. This paradigm makes the epistemological assumption that knowledge is constructed through the interaction of the researcher and the study participants with the outcome being meaningful knowledge (Tavakol and Sandars, 2014a). Ontologically the constructivist worldview assumes that there are multiple realities to explore, and each individual studied provides a unique viewpoint on the area of study. Unlike quantitative research, with its positivist grounding, qualitative research consequently focusses on a process of inductive reasoning to construct theories. In doing so it provides subjective and non-measurable outcomes which have taken time to be accepted within medical education, with sometimes polarized opinions (Ercikan and Roth, 2006).

3.4.2 Ensuring Validity in Qualitative research

With the concerns raised in medicine regarding the validity of qualitative research (a consequence of the positivist paradigm of most underpinning scientific medical knowledge) it is of particular importance that quality and validity of research is considered (Morse, 2006). The classical criteria for measuring validity in research are reliability, validity and objectivity however these sit awkwardly with qualitative research (Flick, 2018). Alternative criteria have been suggested, such as credibility and procedural dependability, however these have been suggested to be new answers to old questions without any new solutions (Miles and Huberman, 1984).

Most authors focus on key steps to be taken within qualitative research in order to ensure that the study meets appropriate requirements for validity which form both part of the process but also must be considered within the formulation of the method of the research itself. Reviewing the considerations of multiple authors allows us to conclude the primary measures of validity as set out in table 2 (Cohen, 2018).

With validity in qualitative research being bound to the process of the research itself these key validity markers will be explored in more detail, and highlighted as such, within the detailed method for the qualitative arm of this study.

Table 2: Primary measures of validity in qualitative research

Descriptive Validity	That the research presented is factually accurate
Interpretive Validity	That the research is able to accurately capture the meaning, interpretation and understanding of the participants
Theoretical Validity	The extent to which the research explains phenomena
Generalisability	That the research theory generated may be helpful to understand other situations
Evaluative Validity	That a judgement is made regarding the subject being researched rather than simply an interpretive framework
Transparency	That the reader of the research is advised of, and can understand, the process of data interpretation

3.4.3 Sampling decisions in Qualitative research

The aim of sampling for qualitative research is to ensure that the results produced can be utilised to support generalisable statements rather than ones applicable only to the test subjects. Unlike quantitative research substantial criteria are utilised rather than formal ones and, in addition, the sampling process is integral to the data collection and subsequent analysis (Flick, 2018).

Theoretical sampling was conceptualised by Glaser and Strauss in the late 1960's. In this approach statistical sampling is not considered but rather the sampling is an iterative process in which the initial sample groups, and the analysis of the data from them, helps inform the ongoing need for further information and targeting of additional individuals for representativeness (Glaser, 1967). This must be done within a clear research question in order to ensure that the data remains focussed and avoid a never-ending tangential exploration of ideas.

The focus in sampling in this way is by relevance of the individuals or groups under study rather than their representativeness. For studies such as this in which data is to be collected from different institutions the question arises regarding the most appropriate breadth of initial sample. This links to data triangulation and the concepts espoused by Denzin who

focused on purposive and systematic selection across groups, time and settings (Denzin, 1989).

This approach is aligned to the concept of purposive sampling which, in the context of this study would be sampling within individual institutions, and amongst clinician types, to get a breadth of experience. Within this overarching purposive sampling approach however there will be a need for pragmatism and recognition that convenience sampling is likely to be required to ensure adequate numbers of contributors to focus groups. Convenience sampling works when the criteria for effective participants can be met, specifically that they have the necessary knowledge and experience to answer the questions posed in the study (primary selection or purposive sampling at a high level) whilst recognising that convenience sampling itself represents a form of secondary selection, i.e. utilising those willing to participate in the study (Flick, 2018).

Sample size also needs to be considered, however it is notable that there is no clear rule for sample size in qualitative research (Braun and Clarke, 2013). The concept of saturation, with its basis in grounded theory research, is useful with the principle being to continue until no new viewpoints are revealed. This however must be considered within practical considerations such as ensuring generalisability, the limits of data analysis within the time available and the issues raised in the previous paragraph. Braun and Clarke summarise as stating that the key is ensuring that the data gathered is sufficient to tell an appropriately rich story.

In summary sampling decisions in qualitative research are based upon the study design and requirements for answering the research questions. Whilst sampling may be driven by the data outcomes it should still be based on robust theoretical underpinnings.

3.4.4 Limitations of Qualitative research

Qualitative research has its limitations which must be acknowledged if the findings are to be appropriately contextualised. The primary disadvantage of qualitative methods are that the generalisability to wider populations cannot be as robustly stated as with a quantitative study as there is no statistical analysis against a fixed hypothesis (Ochieng, 2009).

The other major limitation is linked to the integral nature of the researcher within the research process. The researcher's own skills and biases in the qualitative research

methodology can lead to a reduction in the quality and validity of the output emphasising the need for relevant steps to be taken to prevent against this. These steps include the consideration of deviant responses, ensuring that findings are considered in light of previous findings rather than as stand-alone information, and triangulation through the use of other methods (Anderson, 2010). It is particularly with the latter of these in mind that this study will be undertaken as a mixed methods piece of work. For this study there is also the risk of difficulties where a power dynamic potentially exists between the researcher and the participants as in this study. It will consequently be important that to overcome these potential limitations a robust approach to ensure validity is taken with control of researcher bias where it might affect responses (Salter and Atkins, 2014). This issue is discussed in more detail in chapter 11.

3.4.5 Use of Focus Groups in Qualitative research

A focus group is an interview process undertaken in groups where the researcher supplies a topic but the interaction and discussion between the participants produces the output (Morgan, 1988). The groups bring together appropriate individuals, selected for their characteristics and relevance to the study, in order to discuss a specific area of interest, facilitated by a researcher. It is the interaction between the participants that is important so it is also important that the group can find clear common ground and that the researcher ensures that all participants can effectively contribute to the question set against that common ground (Hydén and Bülow, 2003).

Focus groups have both strengths and drawbacks. There is strength in the ability to gain insights that a single interview may fail to gain through the interaction between the participants and, in addition, it may enable involvement of some participants who might otherwise not wish to undertake a one-to-one interview. Conversely, they may yield less data than the same number of individuals interviewed on a one-to-one basis and the risk of "groupthink" might lead to a loss of the individual voice which might provide an alternative viewpoint.

In this study focus groups were chosen over interviews for pragmatic reasons, as it was simpler to arrange fewer focus groups than more interviews, they would encourage participation through peers, and so that there could be challenge and reflection from

individuals as they considered their own clinical reasoning in light of the answers of others. Homogeneity of subjects (all clinical students) in this study provided some protection against internal hierarchy stifling responses, yet still required that the group was managed, to facilitate all voices and viewpoints to be heard (Kitzinger, 1995).

3.4.6 Undertaking data analysis in qualitative research

Analysis of qualitative data does not have a standardised formula approach as themes are drawn inductively from the collected information with the researcher utilising their knowledge and expertise to produce meaningful conclusions. In interview based qualitative research content analysis is most commonly used in which the interviews are transcribed verbatim into a suitable format before being analysed (Flick, 2018; Kleinheksel *et al.*, 2020). Within the mixed methods convergent study design, the qualitative and quantitative data require comparison for each concept. In order to undertake the qualitative component of the data analysis, it is important to utilise a method which will facilitate this within an appropriate, qualitative evaluation framework. In this case thematic analysis will be utilised to undertake the qualitative data analysis. Given the uniquely flexible nature of the thematic analysis it is important to ensure that the key assumptions within the data evaluation are considered. The thematic analysis approach for this part of the study would be considered a coding reliability model in which predetermined theories provide an initial set of themes against which the coding is set. The codes can be derived from the researchers own knowledge and/or from the evolving themes of the work and do not follow a set methodological requirement (Cohen, 2018), however there is a clear theoretical underpinning assumption to the process (Clarke and Braun, 2013). This is not to say that the initial themes are completely set in stone, and indeed the themes and sub themes may develop through the analysis of the data, however it provides a starting point to facilitate the combining of the two data sets in a mixed method study.

The analysis process involves immersion in the data and creating groups of information which have a common idea or theme. These common ideas are coded to facilitate the interpretation, usually utilising an appropriate software package, such as NVIVO, as is the case in this work. There are methodological debates in coding, inherent in this notion of varying methodological requirement, however in this work the process will align to the six-stage model set out by Braun and Clarke to ensure validity (Clarke and Braun, 2013) as

shown in table 3 This approach ensures a systematic approach and facilitate the triangulation of two methodological processes.

Table 3 – The Six stages of Thematic Analysis

Familiarization	A broad high-level review of the data and noting down of initial thoughts
Generating codes	Methodically label important elements of the data
Generating themes	Draw together different codes into key thematic areas
Reviewing themes	Review the themes generated against the data to ensure they accurately represent it
Defining and naming themes	Define the themes to ensure that they can allow us to make sense of the data
Creating the report	Set out the findings in a coherent narrative

It is intended that the development of the themes will be data driven with generation through the analysis of the early focus group interviews with reliability and validity tested through application to later interviews and relevant revision undertaken. This process can ensure that the coding structure is exhaustive and ensure saturation. This approach will also allow the process and outputs to be quality assured using the approach set out by Braun and Clarke in 2021 and thus further supporting the reliability and validity of this part of the process (Braun and Clarke, 2021).

3.5 Bringing Methodology Together – data interpretation

3.5.1 Challenges of the triangulation concept

Data triangulation in mixed methods research is commonly considered to be a process of confirming the results of a primary methodological approach with a second contrasting one. This view was espoused early in the development of mixed methods research and continues to carry much weight in more recent work (Hesse-Biber, 2010). Concerns have been raised by authors that this approach tends to lean toward the positivist worldview of the quantitative researcher with a nominal qualitative methodological approach being used as a way to confirm the quantitative data (Bryman, 2006). Taking this further a 1989 study by Greene et al found that 25 of 57 studies (44%) pertaining to be of a mixed methods

approach did not undertake data integration during the analysis with either one component acting as a springboard for the other or no attempt being made to bring the data together to form a single combined conclusion (Greene *et al.*, 1989). The authors suggest that these findings indicate that triangulation is not being achieved in many of these studies as there is no convergence of the data sets. The authors see triangulation as one of a number of options for utilising the differing datasets produced from the separate arms of the study. Flick (2017), in discussing data triangulation in mixed methods, notes that there is a conflict that exists within the methodological concept that the combination of quantitative and qualitative data is the only option. This position is at odds with the mixed methods view that it is the research question that drives the methods and that the chosen methods therefore represent a pragmatic decision on methodological purity (Flick, 2017). This study does utilise the typical combination of qualitative and quantitative study types, however, as illustrated above, doing so does not ensure appropriate data integration and conclusion formulation. This integration and interpretation process must be specific to the research question and is discussed in more detail below.

3.5.2 Defining the purpose of the study to drive data interpretation approaches

In their work Greene and colleagues stepped back to look at the purpose of any piece of mixed methods research data interpretation, of which triangulation is one, as shown in table 4. In keeping with the pragmatist worldview these can be applied to this project to help guide data interpretation (Greene *et al.*, 1989).

Table 4: Purpose definitions for data interpretation (from Greene, Caracelli et al. 1989)

Triangulation	To gain convergence and correspondence of results from the different methods.
Complementarity	To gain elaboration and results clarification from one method through the results of the other
Development	Uses the results from one method to inform/develop the other method including sampling and implementation decisions
Initiation	Seeks paradox and contradictions including the redevelopment of questions and results from one method with those of the other
Expansion	Aims to extend the breadth and range of the research question by using different methods for different inquiry components.

Taking the position of the research driving the data interpretation, this study is aligned to the complementarity purpose, with the two approaches seeking to look at the same phenomena but seeking clarity and elaboration from the other (breadth vs depth). The quantitative arm of the study will provide a broad impression of student's approach to clinical reasoning in assessments, but with little depth of understanding, with this depth then explored through the qualitative arm of the study. Triangulation would seem a potentially attractive purpose in this study however the narrative review has shown that the process of reasoning has a number of potential approaches therefore shifting alignment toward the constructivist worldview more than the positivist. Accepting that the results have the potential to give multiple answers it is thus more appropriate to consider the qualitative data as elaborating on the quantitative, thus making a complimentary purpose the most logical approach. Greene and colleagues (1989) consider that understanding the purpose of the research better enables us to understand the methodological principles involved. They suggest that for an effective complimentary purpose mixed methods design we need different methods, exploring the same phenomenon, with equal status and undertaken simultaneously but independently (Greene *et al.*, 1989).

The work of these authors has been built upon and developed by Bryman in 2006 who felt that Greene's work led to some restrictions in the five purposes. Whilst Bryman's work expands on these, it loses simplicity and fidelity in doing so with 18 possible rationales/purposes described (Bryman, 2006). When applied to this study a combination of four of these 18 rationales could be applied to describe the data integration proposed, specifically offset, completeness, process and explanation. These are set out in table 5.

These all fit with the overall complimentary purpose, describing variations, or subcategories, upon this theme. In considering this further this greater granularity does not clearly add additional value to the understanding or methodological approach so the complementarity purpose of Greene et al is that which will be considered here.

Table 5: Definition of the applicable Bryman rationales

Offset	Suggests that the research methods associated with both quantitative and qualitative research have their own strengths and weaknesses and combining them allows the researcher to offset the weaknesses to draw on the strengths of both
Completeness	The notion that the researcher can bring together a more comprehensive account of the area of enquiry if both quantitative and qualitative research are employed
Process	Quantitative research provides an account of structures in social life, but qualitative research provides sense of process
Explanation	One is used to help explain findings generated by the other

3.5.3 Applying these approaches to this project for methodological clarity

Mixed methods research interpretation is driven by the methodology, yet equally the purpose of the data in relation to the research question can be said to influence the way that the components of the methodology are brought together. Within this study the quantitative data will provide a broad descriptive view of reasoning approaches with the qualitative data allowing us to better understand the phenomenon. This fits with the approach of authors who have sought to move mixed methods research away from a position of "quantitative dominance", and thus a positivist leaning in the approach, to one in which the qualitative approach is supported by quantitative data (Hesse-Biber, 2010). Consequently the data will not be directly combined, with no attempt made to quantify the qualitative data. Rather the use of the same questions, derived from the Nominal Group portion of the study, will allow the same themes to be explored and comparisons made between the two datasets to draw conclusion from the work. In the quantitative component this will be the Likert item questions whilst in the qualitative component they will form the structure for initiating components of the focus group discussion.

3.5.4 Method design to support a question driven mixed methods study

The study will utilise a convergent design in which data is collected contemporaneously in both the quantitative and qualitative arms of the study, seeking convergence between the responses in the two arms, and analysing how the results concur or differ (Creswell and Clark, 2017). In doing so, it is also possible to mitigate the weaknesses of the two different

approaches (breadth v depth, depth v breadth), in order to provide a more extensive understanding of the topic.

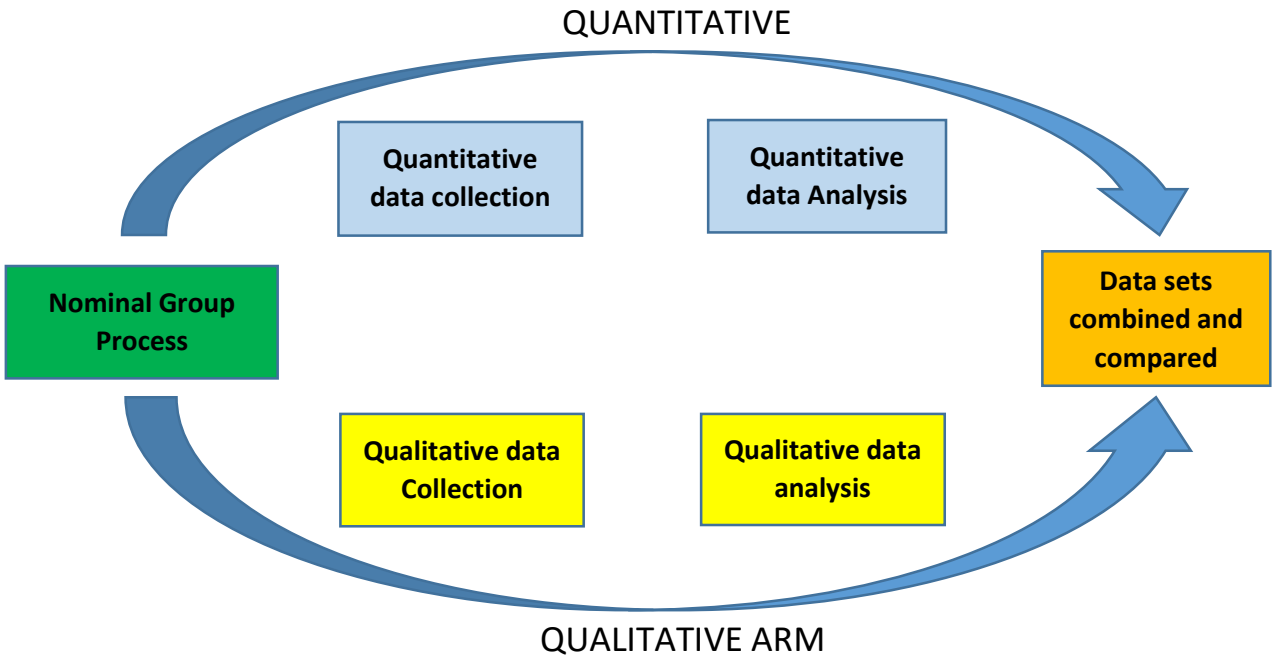


Figure 4: The Convergent Methodology

The convergent methodology involves the collection of data simultaneously in the two elements of the study and places equal importance on them. The data analysis is conducted individually in the two elements before being brought together to draw meaningful conclusions and will be presented as such. Authors describe variants of the convergent design, depending upon the way that the data is collected, analysed and the interaction that occurs between the datasets during collection. This study is expected to use the parallel-databases variant (Creswell and Clark, 2017) with the methodology shown in figure 4.

Strengths of this methodology include its efficiency in data capture, the ability to analyse the data samples using the methods most familiar to the data type (quantitative or qualitative), and the ability to develop understanding of the research question through data developed via both worldviews.

Challenges do exist to this methodology which need to be considered including the difference in sample sizes, the difficulty in merging two very different datasets, and the risk

of the two parts of the study producing differing results. These points will be addressed dynamically as the study progresses and covered in chapter 9 where the convergent results will be discussed.

3.6 Data analysis and integration

Data analysis and interpretation in mixed methods research is tied into the chosen methodology and its process. This part of the methodology involves the analysis of the qualitative and quantitative elements of the research in a way appropriate to their paradigm and data set before integrating the results (Creswell and Clark, 2017).

3.6.1 Common processes in the quantitative and qualitative approach

Whilst the data analysis itself is different between the two research paradigms, the steps are common to both (Creswell and Clark, 2017):

- Exploration – in quantitative research this involves an initial viewing of the data and a descriptive analysis whilst in qualitative research initial thoughts are jotted down during an initial full review. These thoughts start to generate codes for analysis and the development of a code book.
- Analysis of the data – The quantitative data is analysed based upon the hypothesis using an appropriate statistical approach. Qualitative data is reviewed, and themes developed through the coding of small pieces of data.
- Represent the data analysis – Quantitative data is best represented through appropriate graphical plots of the statistical data whilst qualitative data is usually more narratively represented through thematic discussion.
- Interpret the results – In both cases the major findings are summarized with quantitative data being compared against the hypothesis whilst qualitative is considered against the research question and includes the researcher's personal assessment of the data.
- Data validation – This includes both validity and reliability and have been well described previously in this chapter and the approach taken for both in this research will be discussed in chapter 6.

3.6.2 Data Integration

The core purpose of mixed methods research is the additional value that can be gained through the combination of different methodological data, thus enriching our understanding of the research question under investigation. Mixed methods research is designed to improve the quality of our inferential conclusions drawn from the strengths of both approaches (Teddlie, 2009). Previous authors have noted the lack of literature focussed on the integration of mixed methods data with many studies consequently failing to achieve a "sum greater than the parts" within the research output (Bryman, 2006). Recognising that the methodology drives the interpretation the four key steps (data integration through the intent of the integration, the data analysis procedures, the representation of the results and the interpretation of the results) are defined directly by the approach taken, in this case the convergent approach (Creswell and Clark, 2017).

Recognising that the intent of the study is to develop an expanded understanding of the phenomenon under investigation it is important that both arms gather data on similar concepts to ensure validity in data integration. This strengthens the rationale for the use of the Nominal Group Approach to define the questions on which the two arms of the research will be based (see below). With both arms looking at the same concepts but in different ways the data can be integrated into a narrative discussion around quantitative statistics and qualitative descriptions allowing clarity of convergence or divergence of the outputs. It is important to consider inconsistencies in the data between the two arms of the study. With a study such as this there is considerable risk of this occurring, with the recognition that a single truth may not be present given the subject matter, however this must be explored, recognised and appropriately contextualised (Slonim-Nevo and Nevo, 2009).

This approach to the combined data analysis will permit an abductive approach to generation of conclusions in relation to the research topic with a "best explanation" approach utilised to address the areas of interest.

3.7 Facilitating validity through use of Consensus Group Methods

Validity in a study such as this can be aided through ensuring that the same concepts are being explored in both arms of the study. This study will utilise a consensus group method

to define the key questions to be asked in both the quantitative and qualitative parts of the study.

3.7.1 Consensus group methodological approaches

There are three main methodologies in consensus group studies, Delphi, Nominal group technique and the RAND appropriateness method. These methods do not easily fit into the natural distinctions of qualitative or quantitative research and some consider them to have elements of both (Stewart, 2001). Table 6 sets out the distinctions between the three approaches (McMillan *et al.*, 2016). It is for this reason they are being considered separately within this chapter. These methods do not have a clear philosophical worldview against which to be considered, with many versions, particularly of the Delphi method, existing with inconsistent methodology descriptions within published reports (Humphrey-Murto *et al.*, 2017b).

Table 6: A comparison of consensus methods

Nominal Group	Highly structured face to face group interaction focussed on group generated ideas
Delphi	Highly structured remote group interaction using repeated questionnaires based on researcher ideas
RAND	Initial review of a detailed literature review and then similar to Delphi but with a face to face discussion after the first round

These approaches are commonly used across a range of research areas including healthcare. Their use is designed to provide an appropriately systematic approach to gaining expert consensus in situations where evidence is lacking (Murphy *et al.*, 1998). Expert opinion sits at the bottom of the levels of evidence (Guyatt *et al.*, 2008) yet, if opinion is gathered in a structured and appropriate way it can provide useful outputs in areas which are otherwise difficult to formalise through an alternative research methodology.

3.7.2 Ensuring rigour in consensus group techniques

Whilst lacking in clear philosophical grounding it is still crucial to ensure appropriate rigor in consensus group techniques. It is important therefore that a series of specific steps are followed shown in table 7 below (Humphrey-Murto *et al.*, 2017b).

Table 7: Key steps to ensure methodological rigor in consensus group techniques (from Humphrey-Murto, Varpio et al 2017).

1. Define the purpose or objective of the study
2. Outline each step in the process: if modifications were made, provide a rationale for the choices made
3. Describe the selection and preparation of the scientific evidence for the participants
4. Describe how items were selected for inclusion in the initial questionnaire: describe the process in sufficient detail
5. Describe how the participants were selected and their qualifications: If NGT is used describe the facilitators credentials
6. Describe the number of rounds planned and/or criteria for terminating the process
7. Clearly define how consensus was defined
8. Report response rates and results after each round
9. Describe the feedback provided after each round
10. Describe how anonymity was maintained
11. Address potential methodological issues in the discussion

3.7.3 Nominal Group Technique

The Nominal Group Technique (NGT) is a consensus group technique designed to facilitate a rapid outcome through a single meeting. It consists of a group of experts undertaking an idea generation, discussion and ranking process to agree the outcomes and their importance (Jones and Hunter, 1995). The process is illustrated in Figure 5. NGT is particularly suited to studies where the exploration of ideas is key and where the ability to get individuals to participate in multiple rounds may be difficult which forms the core of the Delphi process.

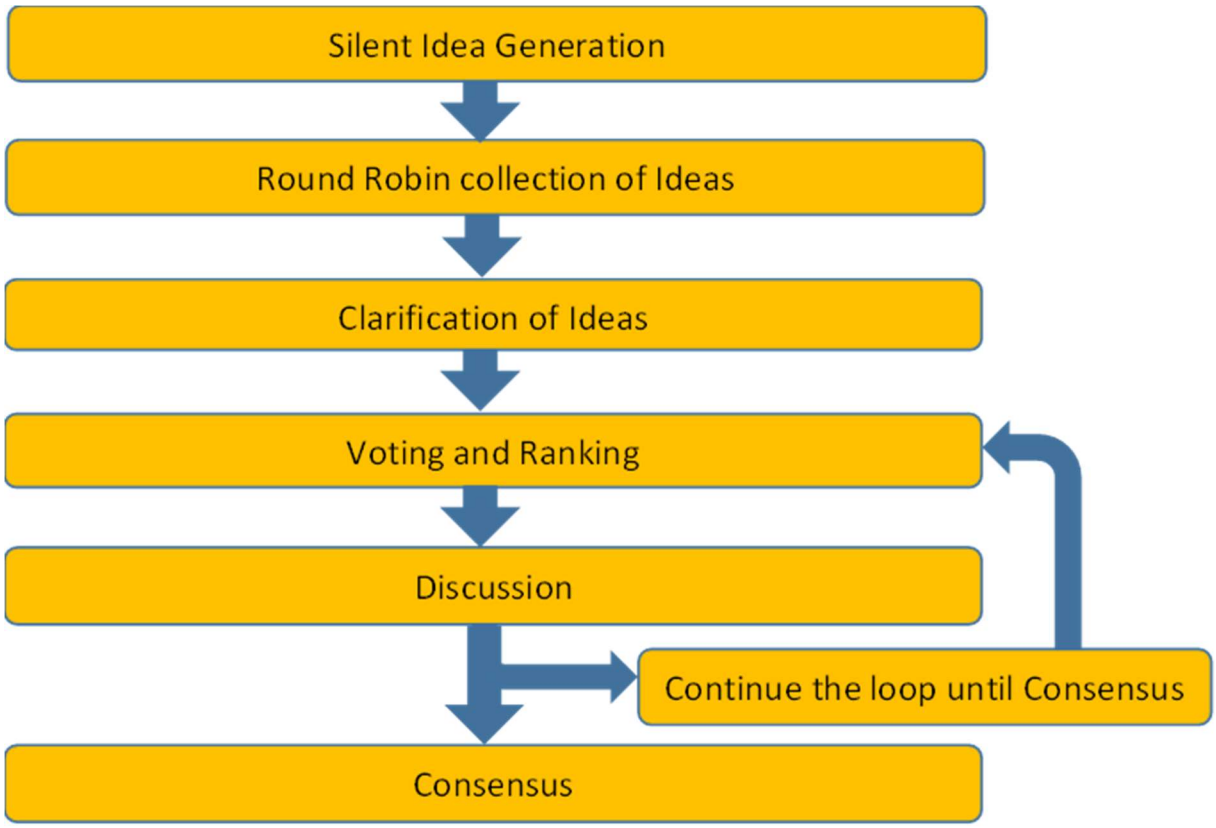


Figure 5: Key steps in the Nominal Group Technique (from McMillan, King et al. 2016)

There has been suggestion that NGT is slightly superior to Delphi in providing accurate outcomes when compared to a simple meeting (Rowe *et al.*, 1991). Maintaining quality and validity in this method is through the use of anonymity during the initial idea generation stage and the later rating stage. Utilising the NGT approach also reduces the potential risk of bias that can occur in the Delphi process due to the initial idea generation from the lead researcher. The process should follow a clear structure (figure 5) (McMillan *et al.*, 2016) and ensure that the concerns raised by other authors are addressed in order to ensure that this claim of rigour can be justified (Humphrey-Murto *et al.*, 2017b). The details of the methodological approach will be covered in chapter 6.

3.8 Summary

Through this chapter the study methodology and its theoretical basis and rationale have been discussed. It is important to note that mixed methods research is dynamic and therefore it is common for changes to be required during the process. These will be

reflected within the detailed methods should any such changes occur as the study progresses. The elements of the convergent design process and their detailed methods will be discussed in subsequent chapters along with the data interpretation at each stage.

Chapter 4 – The Nominal Group Technique – Generating questions

4.1 Introduction and Chapter Summary

The Nominal Group Technique (NGT) has been considered in the context of other consensus methods in chapter three, setting out its benefits and the processes required to ensure it works as intended. As previously stated, its use in this context is to explore the ideas that a group of experts have within a defined time scale.

This chapter will set out the benefits of the Nominal Group Approach before revisiting the process itself. The chapter will then go on to describe the NGT in this study, including recruitment and preparation, before taking the reader through the process undertaken and describing how the steps in figure 5 were undertaken. There will be a clear description of the online process used as this is different to the usual approach to NGT which is more commonly done on a face-to-face basis. Finally the chapter will set out the outcomes of the process which will then become the core question set for both the quantitative and qualitative arm of the study.

4.2 The NGT process in this research

In this research study the NGT approach is being used to set the key questions that will form the basis of both the Likert data items within the quantitative arm of the study and the questions for the semi-structured focus group interviews within the qualitative arm. The key themes coming out of this process will also form the overarching coding framework for the thematic analysis of the focus group data. With a topic such as clinical reasoning the narrative review (chapter 2) has already demonstrated the significant plurality of ideas and concepts that exist, and it is therefore important to define the best questions through the involvement of experts in the field. The primary purpose of this step is therefore to set questions to better understand the clinical reasoning process used by students completing an assessment, specifically SBA and OSCE. The concept framework developed within the narrative review will be used to serve the basis of the NGT process. This step in the research represents a way to develop the key questions through the broader understanding of other

experts in the field (Van de Ven and Delbecq, 1972) rather than reliance on a single researcher. It was decided to target the development of ten questions, spread across the full range of the concept framework, as this represented a pragmatic number for student completion and avoids duplication of specific areas. It also ensured that the focus groups would not risk becoming too "question heavy" as the same questions would form the basis of the discussions within these.

The nominal group process follows a series of clearly defined steps already outlined in the methodology. These are described in more detail in Figure 5, already seen in the methodology chapter. Prior to the actual process of the NGT it is important to ensure appropriate selection of participants, the selection of, and provision of the relevant scientific information to facilitate the process, and relevant ethics and process approvals (Humphrey-Murto *et al.*, 2017a). These will be considered below with each stage described to enable the method taken to be followed.

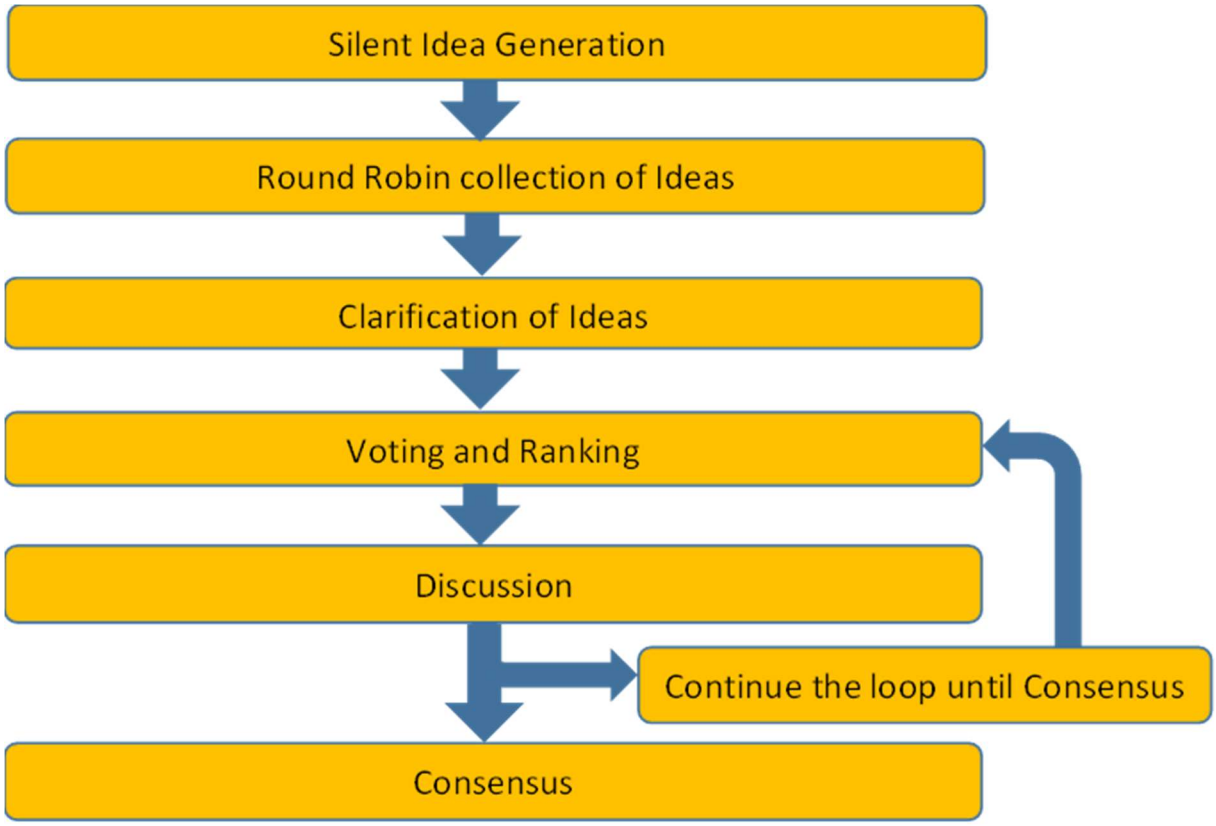


Figure 5: Key steps in the Nominal Group Technique (from McMillan, King et al. 2016)

Ethics approval was obtained for this section of the work via the University 1 ethics approval process application reference 043363 included as appendix 3.

4.2.1 Developing the expert group

The first step in the NGT is defining the expert group who will participate in the process. This constitutes the need to have an appropriate number in order to get a meaningful output whilst recognising that a balance of views on the subject is more important than excessive numbers. Work looking at this issue has suggested that seven members provides an optimal number for this process, although two to 14 have been used (McMillan *et al.*, 2014). In this case expertise was sought from the UK Clinical Reasoning in Medical Education (CReME) group as experts in the field and of whom the researcher is a member. A general invite for expressions of interest was sent out with the group selected from those responding to this, seeking a breadth of participant geography and institutions to reduce potential sources of bias. Additionally further potential participants were identified through colleagues who had an interest in, and already taught on, the topic of clinical reasoning locally and invites sent.

The recruited expert group consisted of six members, including myself, and this number sat within the range noted by McMillan et al. One of the project supervisors acted as an independent facilitator ensuring that I was able to participate. This was also important to provide a review of the process and monitor for potential risk of bias due to my own involvement.

4.2.2 Provision of Scientific Evidence and information for the Participants

The primary source of scientific evidence for the participants was the narrative review of chapter 2. This was provided to participants in a condensed format, with the focus on the development of the concept framework of clinical reasoning (chapter 2, 2.4 to 2.7 inclusive), which forms the basis of the key areas to be explored within the qualitative and quantitative arms of the study. Additional information was provided for participants: an outline of the purpose of the overall research, the purpose of the NGT part of the research, a narrative description of the NGT process and the steps involved. This ensured that all could be fully informed prior to the day with the aim of maximising the efficiency of the process. The Information for participants and process detail documents are included as appendices four and five. The process information (appendix five section three) provided examples of output

types, alongside how they were to be utilised, in order to ensure that the final results from the process were fit for purpose in the subsequent stages of the process. Participants were provided with a copy of the ethics approval. Consent forms were sent out prior to the day using a google form to gather informed consent. This was checked and completed on the day itself.

4.2.3 Developing the NGT information gathering tools

Due to running the process remotely, it was necessary to utilise collaborative online working processes to effectively gather the information from the participants. This was done using two google sheets (Google, 2022b), one for the initial question capture and a second for the ranking process.

On the initial google sheet each participant had a separate page in which to enter their initial questions during the silent question generation phase. This ensured anonymity and that each participant generated questions independent of the others. These sheets auto populated a single page which could then be used for the question review (round robin collection of ideas and clarification). Following the question review process the remaining questions were automatically drawn through into the second ranking spreadsheet both onto a master ranking page and a separate page for each individual participant to undertake their ranking (ranking of ideas). This ensured that each participant could perform the ranking individually without influence from seeing the other members populating their own scores. These rankings fed back into the master page with rank calculations completed automatically.

The use of spreadsheet software online was facilitative and ensured the process could be completed simultaneously by multiple remote participants. It also simplified the process of ordering the ranking at the end of the process. The overall process is described visually in Figure 6. Screenshots from the process are included as appendix 6.

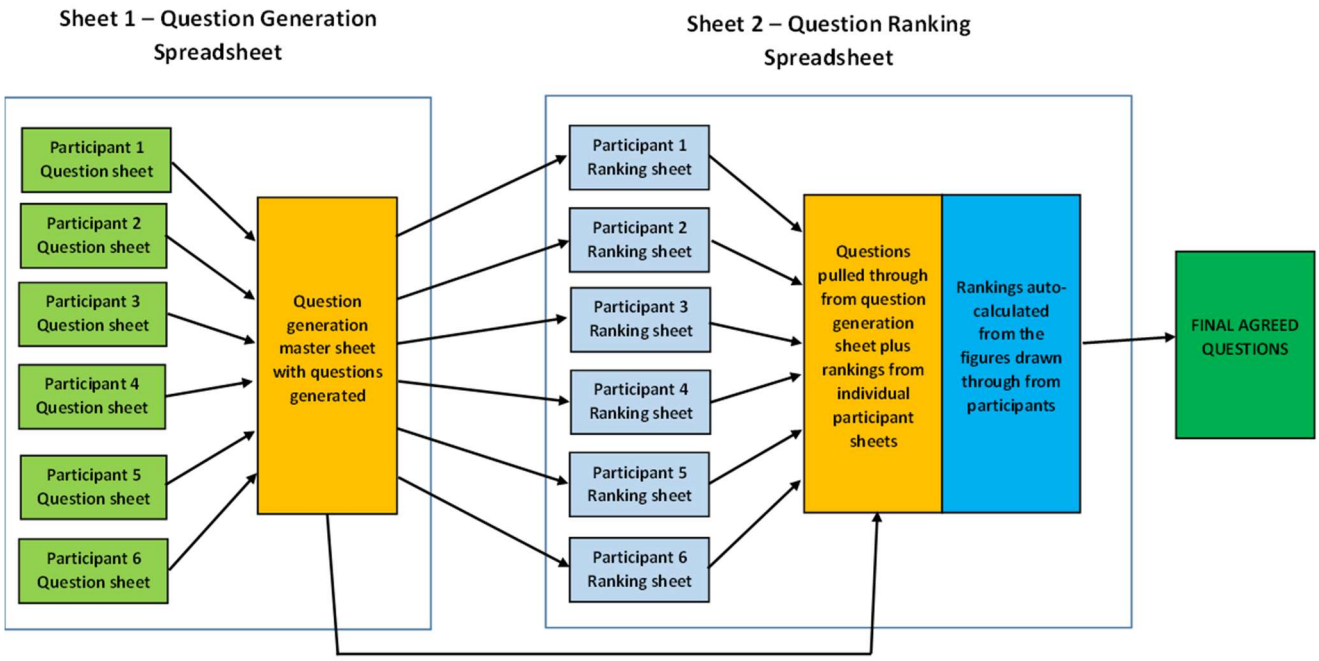


Figure 6: Diagrammatic representation of the spreadsheet process to undertake the nominal group process

4.2.4 Initial participant briefing

The NGT process was run remotely using Google Meet (Google, 2022a) video conferencing software. Due to emergency clinical commitments, one participant withdrew from the process the day before and another was running late for the session, thus four individuals took part in the initial question generation stage of the process. This number still sits within the range noted as having been used before in NGT processes (McMillan *et al.*, 2016). Prior to commencing the silent question generation stage the group were briefed regarding the overall purpose of the research project to ensure that the questions generated were correctly framed against this. Participants were also briefed further on the NGT process, the correct utilisation of the spreadsheets, and timings planned for the session. Participants then were able to work individually to generate questions.

4.2.5 Question Generation Stage

In the initial individual silent question generation stage (utilising the question generation spreadsheet) 68 questions were put forward. Initially each question author read through their questions, to allow the identification of any possible duplication. Whilst undertaking this process the previously delayed colleague was able to join us leading to five reviewers. It

was clear that there was significant duplication between the questions, so to facilitate a reduction in the numbers to take forward to ranking, each question was tagged with a label to facilitate the grouping of questions for review (Table 8). These tags were agreed by the participants within the meeting and generated as each question was considered. These topic tags also formed the initial coding framework for the thematic analysis. It is interesting to note during this process that, even amongst this group, representing expertise in decision-making, some differing of understanding of certain terms was evident.

Table 8: Topic tags used to group questions for comparison.

Abductive reasoning	Biases	Deductive reasoning	Early hypothesis generation
General	Heuristics	Metacognition	Patient centred context
Reasoning	Red flags	Satisficing	Student centred context

Following this process, the groups of questions were reviewed by all participants and those that were either very closely worded or felt to fall beyond the parameters of the research question, were removed. In total 27 questions were removed with 41 taken forward to the question ranking stage of the process.

4.2.6 Question ranking stage

The 41 questions were automatically pulled from the question generation spreadsheet to the question ranking spreadsheet. All five group members took part in this stage, designed to produce the final agreed questions. Each group member individually ranked the questions with these automatically populating a master sheet which then calculated the total rankings. It was decided that each individual would rank their top twenty questions (with 1 being the question deemed most important and 20 the least), with any questions outside the top 20 being given a nominal rank of 30. This approach was taken as the group felt that outside the top 20 decisions regarding the relative merit of any one single question over another became increasingly arbitrary. Top 20 was chosen to allow for variation in views but recognising that, as the target was ten questions, ranking beyond 20 was unlikely

to be helpful. This process led to a ranked list of questions from which it was felt the top ten should initially be reviewed.

4.2.7 Final Agreement of Output

The top ten questions were reviewed against the overall aim of the research project, particularly as to whether an appropriate spread of questions, covering the concept framework, had been developed.

Of the top ten it was felt that two questions were very similar and should be combined into a single question, specifically:

> *"I look for recognisable patterns within the presentation / scenario"*
>
> *"I find it easier to come to a decision if I recognise the case as similar to one I have seen in practice"*

These were combined into:

> *"I look for recognisable patterns within the question to generate my answer"*

As this now left nine questions it was discussed and agreed to review the following ten questions to see if any questions further added to the stated aim of the research project in terms of covering the clinical reasoning process as set out in the concept framework. The 15th ranked question was discussed by the group and agreed as adding value to exploration of the tendency toward deductive reasoning.

The final agreed questions are shown in table 9 along with the primary element of the clinical reasoning process being tested by each one agreed by the group, framed against the research question which pertains to the process in assessment. These are listed in their forms for both the quantitative and qualitative arms of the study and in a final order to mirror the concept framework. Decision-making was the term used most commonly by the group formulating the questions and it was felt appropriate to use this term in the questions, rather than clinical reasoning, to aid the understanding of the participants who were not being provided with any specific teaching on the topic prior to their participation.

These final questions were compiled and sent out to group members for a final consensus agreement and confirmation that these accurately represented the groups work.

It is important to note that the questions developed for the two arms of the study differ slightly due to their requirements. Those for the quantitative arm need to enable answers on a linear scale whilst those for the focus groups act as starting point from which conversation can develop and clarification can be sought.

Table 9: Final Agreed Questions for the mixed methods study

	Final Quantitative Question (Likert scale)	Final Qualitative question (Focus Group)	Focus of Question
1	How I feel in the assessment affects the way I make decisions in questions (never – always)	How does the way you feel in an assessment affect the way that you make decisions in questions	Student centred context
2	The patient demographics have an impact on my decision-making (never – always)	How do the patient demographics have an impact on your decision-making	Patient centred context
3	The clinical setting of the question has an impact on my decision-making (never – always)	How does the clinical setting of the question have an impact on your decision-making	Patient centred context
4	I usually come up with hypotheses immediately then change them as the question develops (never – always)	Do you usually come up with hypotheses immediately then change them as the question develops	Early Hypothesis generation
5	I usually generate a list of possible answers and eliminate them one by one (never – always)	Do you usually generate a list of possible answers and eliminate them one by one	Deductive reasoning
6	I try and keep an open mind until I have formulated an answer (never – always)	Do you try and keep an open mind until you have formulated an answer	Deductive reasoning
7	I look for recognisable patterns within the question to generate my answer (never – always)	Do you look for recognisable patterns within the question to generate your answers	Pattern Recognition
8	I tend to come to an answer but am not sure how I got there (never – always)	Do you tend to come to an answer but find that you are not sure how you got there	Inductive reasoning/Pattern recognition
9	I aim for a "good-enough" answer rather than a perfect one (never – always)	Do you tend to aim for a "good-enough" answer rather than a perfect one	Satisficing/Bounded Rationality
10	I acknowledge and try to avoid cognitive biases in my thinking (never – always)	Do you acknowledge and try to avoid cognitive biases in your thinking	Biases

4.3 Discussion

This NGT question generation session followed the process as described by McMillan et al. and thus was undertaken appropriately. The use of this process was designed to improve the validity of the overall study by ensuring an expert group generated the questions to be used. In that respect the aims and objectives of this portion of the study were met.

It is notable that none of the ten questions has been directly linked to metacognition however metacognition is the ability to be aware of, and control, one's own thought processes. Consequently all of the questions relate to elements of metacognition in so far as students are reflecting upon and articulating the process that they undertake at different stages of the clinical reasoning event.

Potential limitations were the number of individuals involved in the initial question generation process and the identified challenge, even amongst experts, of a uniform understanding of the language involved in decision making. It was agreed by the fifth group member that the generated questions covered a sufficiently broad range and so additional question generation by them was unnecessary. As they contributed fully to the discussion and ranking it was felt their contribution, albeit part way through the process, supported rather than detracted from the process outcomes.

4.4 Summary

The Nominal Group process in this research project was designed to produce a set of questions to be utilised in the two arms of the mixed methods study and in that respect achieved its aim. These questions are representative of the full concept framework of clinical reasoning previously generated in the narrative review. In the two forms presented in table 9 they will form the questions for the two arms of the study.

The use of the NGT process to generate these questions reduces the risk of researcher bias in the question formulation and thus improves the overall study validity.

Chapter 5 –Pilot Study

5.1 Introduction and Chapter Summary

This chapter sets out the methods used for developing the intervention resources and decisions for the primary study (such as sample size) which informed the approach to analysing the results. The pilot was also undertaken to test for feasibility. The study design is a mixed methods study utilising a convergent approach.

The chapter will discuss the way the resources, SBA and OSCE questions, were developed for use in the study as these were initially tested within the pilot study. This includes a discussion on question quality assurance. The chapter then considers sample size and the approach to the statistical assessment of the results. This needs to be tested as part of the pilot to ensure validity of the primary study.

Following this consideration is given to the purpose of the pilot and the questions to be answered before going on to discuss how the pilot was conducted in this study in detail. This includes discussion around the methods used and the analysis processes for both study arms. Detailed methods, including aspects such as recruitment and sampling for the qualitative component, will be set out in the methods chapter.

The Pilot itself was designed as a test of feasibility of the process, rather than being designed to test the hypothesis. It will therefore conclude with a consideration of the success in relation to the required areas for testing under the headings of process and resources, management and scientific.

It is worth noting that if considered appropriate the collected data can be used as part of the wider data set unless significant feasibility issues become evident during the data collection process. In this study the data was not reused.

5.2 Developing Intervention Resources

This study relies on the use of question resources, Single Best Answer (SBA) and Objective Structured Clinical Examination (OSCE), as its primary testing approach to stimulate student thinking around the way they undertake clinical reasoning. Well written questions are

imperative to maximise the chances of getting appropriate responses from the students in respect to the research question being asked.

5.2.1 Developing the single best answer questions

Single best answer questions with five possible answers remain one of the commonest ways to test clinical knowledge and are widely used in the United Kingdom and internationally, highlighted through their use for the UK medical licensing examination (GMC, 2024). Despite challenges to their use, with respect to testing of clinical reasoning, evidence suggests that the writing of good questions will permit us to effectively test this (Palmer and Devitt, 2007). Whilst some authors have questioned the value of five answers, with the potential for a number of answers to be implausible due to the difficulty in writing them (Kilgour and Tayyaba, 2015), this remains the current standard question type in use. In the United Kingdom the guidelines set out by the Medical Schools Assessment Alliance (MSCAA) regarding quality within question writing are those that should be aspired to and they produce detailed guidance to aid question writers (MSCAA, 2019).

5.2.2 Decision-making on the number of SBA questions to develop

In order for the SBA test in this study to be meaningful in stimulating student clinical reasoning, sufficient questions needed to be undertaken. Conversely it was felt that producing an excessive number of questions risked participants becoming mentally fatigued by the time that they came to complete the study questions with the potential for a reduction in the quality of the answers obtained in the Likert data items. Whilst evidence suggests limited performance change with fatigue there is evidence of subjective fatigue which would be best avoided (Jensen *et al.*, 2013). As there was no need to use the questions as a pass/fail assessment, standard setting was not required. This also means that the use of reliability statistics to determine the number of questions could not be used to add value in answering the research question. Pragmatically it was felt that the overall time for students to spend on the SBA questions should be no longer than one hour, in order to ensure that the study was effective, yet not overly demanding on their time.

The MSCAA was approached to establish what they considered to be standard for the length of time to be allocated for a single SBA question, however no response was received. A guide for clinicians on writing SBA questions states that students take sixty to ninety

seconds to answer an SBA question without offering any evidence to support this (Walsh *et al.*, 2017). A study looking at questions being answered electronically, with timing of the completion of the questions measured, showed that an SBA question took an average of 75 seconds to be answered (Huwendiek *et al.*, 2017). Other than this I was unable to find any evidence base for the time allocated to a specific SBA question. The decision was therefore taken to allocate 75 seconds per SBA when calculating the number of questions to use as this fell within the 60-90 second time scale and was in line with the findings of Huwendiek in 2017.

With the need to ensure that adequate time was provided for students to answer the test questions, it was decided that 40 SBA questions would be developed, which should take approximately 50 minutes to be completed. This left ten minutes for the research questionnaire to be answered, meeting the proposed one-hour timescale.

5.2.3 Decision-making on the number of OSCE stations to develop

An OSCE examination consists of a circuit of multiple stations which can vary in number depending upon the way it is to be used in decision-making, however most summative clinical examinations are between 10 and 14 stations. The OSCE stations generated for this study were proposed to be utilised during formative OSCEs, with potentially less stations. It was therefore likely that individual institutions would have very specific requirements about the overall content of the OSCE whilst recognising that clinical reasoning would be required in stations throughout the circuit. It was therefore considered appropriate to write two standardised stations that could be inserted into the formative OSCE of any organisation as part of a larger circuit. The majority of OSCE circuits have a rest station and thus it was planned that the OSCE stations would be run back-to-back before a rest station where students could complete the research questionnaire on a tablet computer. The use of the two stations in this way permitted an evaluation of reasoning using two standardised stations to focus the student prior to answering the questions. This approach sought to provide reduction in potential bias that might occur if no standardisation was in place in the different circuits used for the intervention. The intention is that this would therefore improve overall validity.

5.2.4 Question writing and quality assurance

I am an experienced question writer having developed SBA and OSCE questions for both physician associate and medical student examinations, and for postgraduate royal colleges. All questions were therefore written from scratch by me with detailed records kept ensuring that there was suitable breadth of clinical topic and content.

In order to adequately address the definition of clinical reasoning used in this study, SBA questions were written requiring decision-making regarding both diagnoses and interventions as these latter questions require a two-step process by participants. A two-step process is one in which it is required to first make a diagnostic decision to enable the most appropriate therapeutic one to then be made thus representing both elements of clinical reasoning process. A single SBA was written that had the location context changed (Emergency Department vs General Practice) to see if this altered decision-making as this would be interesting in the overall consideration of context effects and potentially open up new research avenues in the future. The participants would be unaware of there being two versions when completing the study questions.

Two OSCE stations were written on clinical topics that are commonly a diagnostic and therapeutic challenge (abdominal pain and "tired all the time") and require significant clinical reasoning due to the nature of their presentation.

When developing questions for use in examinations best practice involves the use of peer review of questions to ensure that they meet the guidelines and are fit for purpose (Walsh *et al.*, 2017). This quality assurance process was undertaken for these questions to ensure that they were externally reviewed and validated. A group of five clinical academics from University 1, who were highly experienced in quality assurance processes and question evaluation, were invited to take part and undertook a detailed review of the question materials in a structured process.

A sample of the questions used are included as Appendix 7 (SBA questions) and Appendix 8 (OSCE questions).

5.3 Determining Sample size and statistical analysis approach

Effective interpretation of quantitative data relies upon an appropriately determined sample size (Jones *et al.*, 2003). Statistical advice was sought from subject matter experts within the home institution. Likert item data is not normally distributed and therefore non-parametric statistical tests might be considered more appropriate. However there is debate in the literature as to whether Likert item data can be analysed with parametric statistical methods. Likert scales are termed ordinal data and therefore the assumptions of parametric tests aren't met, however when the scale is from 1 – 10 (never to strongly agree, as in this case) then it is argued that it can be treated as interval data (equal distance between data points) and therefore parametric test assumptions can be met (de Winter and Dodou, 2010). The central limit theorem also applies in this situation which states that with a sample of sufficient size, considered by statisticians to be n > 30, the sample mean has an approximately normal distribution, regardless of the distribution pattern of the original data (Rumsey, 2003). Guidance from the Professor of Medical Statistics was to undertake parametric analysis as the primary approach whilst using non-parametric analysis as sensitivity testing through bootstrapping which is supported in the literature by others (Hall and Hart, 1990). Despite this assumption for data interpretation, in order to calculate an ideal sample size both parametric and non-parametric approaches were applied to agree a figure.

Of note the overall population was considered to be one year of graduating Medical and Physician Associate students across the UK. This equates to a total of around 10,000 per annum based on national available data and known expansion in the sector (2020 8750 medical graduates, 2021/22 825 Physician Associate graduates)

Non-Parametric (non-continuous data) sample size calculation can be undertaken with Slovin's formula: $n=N/(1+Ne^2)$ (Black, 1999)

n=sample size, N=total population to be sampled and e=error level (confidence)

This leads to $n=10,000/(1+10,000x0.05^2)$ giving a sample size of 385.

Parametric (continuous data) sample size calculation is undertaken with the following formula : $n=(Z^2 \times StdDev \times (1-StdDev))/e^2$ (Rumsey, 2003)

n=sample size, Z=standard score, StdDev = estimation of deviation from the mean, e=desired confidence interval.

The Z-score for 95% confidence interval is 1.96, Standard deviation is set at 0.5 as this gives the most generous sample size (ensuring it is not too small), 3 is 0.05 representing 95% confidence.

This leads to $n=(1.96^2 x0.5x0.5)/0.05^2$ giving a sample size of 384.

The similarity of these figures therefore permits conclusion that a maximum ideal sample size of 385 is sufficient to draw conclusion for the population under study.

The assumptions made above will also be considered below in relation to plans to compare two samples within the results.

5.4 The Pilot Study

Prior to large scale data gathering it was important to undertake a pilot of both the quantitative and qualitative arms of the study. A pilot study is a test of the methods and procedures to be used on a larger scale, enabling feasibility testing of both the process and analysis approaches, as well as potential identification of effects of interest within the larger or future studies (Thabane *et al.*, 2010). Van Trijlingen and colleagues have grouped the reasons for undertaking a pilot study into four main headings – Process, Resources, Management and Scientific (Van Teijlingen *et al.*, 2001; Van Teijlingen and Hundley, 2002). Table 10 shows the primary reasons for this pilot study.

The focus of this pilot study was feasibility, not statistical significance. On this basis, the questions to be answered (as set out in table 10) to evaluate the success, or otherwise, of the pilot are considered rather than any specific statistically meaningful output. The discussion of the results of this Pilot includes specific reference as to how it informed the planning of the primary study.

In order to potentially permit the use of the pilot data within the primary study the key features of the primary study were also part of the pilot; specifically participants who were part of the target demographic, the resources that were intended to be used for data capture, and the tools that were used for data analysis of the main study. It is likely however that for the qualitative arm of the main study re-analysis may be required as the sub-coding

framework develops in order to ensure that all nuances of the data set are captured effectively.

Table 10: Questions to be answered by undertaking this pilot study

Process	• Is recruitment feasible based upon the student offering?
Resources	• Are the recruitment criteria fit for purpose? • Can the data collection tools capture the required data? • What is the time requirement for completion of the data capture in both arms of the study
Management	• Is the data capture tool robust and will it sufficiently manage the data? • Does the necessary recording media work effectively and as expected? • Are all potential errors in the data capture tool removed? • What capacity management needs to be considered (physical space for study arm delivery)? • Is the researcher able to effectively carry out their required role, particularly within the qualitative arm of the study?
Scientific	• Is the data received in a format to enable appropriate evaluation? • Can the data be extracted to a format for analysis, and can the required analysis be performed with the planned tools? • Can the quantitative data inform the analysis approach? • Does the qualitative data lend itself to the proposed methodological approach? • Is the combining of the data feasible based upon the analysis from the pilot data?

This section will set out the details of the pilot based on the checklist developed by Thabane et al (2010) based on their observation of a lack of appropriate rigor in the way that pilot studies are reported (Thabane *et al.*, 2010).

5.5 Pilot study methods

All final year physician associate students from university 1, who had recently completed their final exams, were invited to participate. Being at this stage in their studies meant that they were part of the study target group adding validity to the pilot. Seventeen participants were recruited for the quantitative component and seven participated in the qualitative component. As all students were in their final year, as per the main study, baseline knowledge testing was not undertaken. Differences in individual knowledge at the same stage in their studies, and their potential effects, was out with the intent of this study.

5.5.1 Quantitative Component

It was decided that the SBA would be the pilot quantitative component as a final check that the questions performed well. It was not considered feasible to undertake an OSCE within the timescale available. In addition, as the same research questions are utilised in the research regardless of the exact assessment type used the data derived would still be suitable for analysis and evaluation within the pilot context. It was recognised that the actual delivery of the OSCE differs from the SBA and thus the first OSCE in which the intervention was utilised also acted as an opportunity to test feasibility.

A quiet room was provided for the participants all of whom were requested to bring their own laptop or tablet computer in order to facilitate answering the questions under exam style conditions. Due to having left the region on completion of their course, two participants joined remotely. In their case they kept their web cameras on throughout the SBA answering process for assurance that they were answering accurately and not utilising external materials to aid their answers with its consequent potential impact on the answers given for the study questions.

When all participants were in place, they were randomly allocated to the two question sets (these are identical other than the one question in which we sought to evaluate if question context made a difference). Consent was gained on the google form itself (as were the answers to the SBA questions).

On completion of the SBA questions, and the research questions, participants were able to see detailed explanations of the question answers to aid their own learning. This was felt to be of significant added value for those participating in the study.

Following the completion of the pilot, the data was downloaded into Microsoft Excel (Microsoft, 2023) and into SPSS (IBM, 2021) for analysis.

5.5.2 Qualitative component

The Qualitative component of the pilot consisted of a focus group, conducted by me, with seven participants drawn from the same group that participated in the quantitative component of the pilot study. As described in previous chapters the qualitative component utilised the same questions as the quantitative component as the starting point for the dialogue.

The group discussion was captured in two ways for the purposes of the research. Firstly Kaltura capture software was used via an encrypted laptop with a USB boundary microphone recording audio and the computer visuals allowing confirmation of which participant was talking at any point as well as audio transcript capture. As a backup to this, a recording was made on an encrypted phone recorder. This was intended purely in case of failure of the computer capture approach.

The researcher acted as facilitator for the group, encouraging contribution and ensuring that discussion became self-sustaining.

Following completion of the group, the recording was initially transcribed utilising the machine transcription process within Kaltura (Kaltura, 2020), which is estimated to be 80-90% accurate by the software manufacturer. This produced a continual transcript without paragraphing or division of individuals. This transcript was then transferred into a word document and the researcher went through putting it into a formal "script" format, correcting machine transcription errors, ensuring narrative accuracy and identifying each individual speaking to facilitate analysis.

This transcript was then loaded into NViVO20 (Lumivero, 2022) for qualitative analysis. The initial qualitative analysis coding frame used the topic tags as set out in chapter four, table 8 (p69) for evaluation of the questions generated in the NGT process.

5.6 Results of Pilot Study

5.6.1 Recruitment

Recruitment to the pilot study was through students from university 1. Seventeen students participated. In this case the group were preparing for another external examination and thus the opportunity to undertake some practice questions was warmly welcomed. The group feedback was that the opportunity to review the answers to these questions – in particular the reasons as to why each particular answer was correct, or otherwise, added significant value.

The feasibility of gathering sufficient data from the target population was considered to be relatively straightforward provided the data collection is positioned at such a time as to be beneficial to the participants in relation to final examinations.

5.6.2 Baseline Data – Quantitative Component

The baseline data from the quantitative component was downloaded from the Google form in .csv file format, permitting analysis in Google Sheets.

Primary statistical analysis was undertaken for each specific research question utilising the in-built formulae within the software package as set out in table 11. It is important to recognise that the analysis is of independent Likert data items, rather than the overall scale (a descriptive term to indicate a series of items) as each is designed to answer a specific question.

Table 11: Primary statistical analyses of the quantitative data

Median	Indicates the central score of the data - gives an idea of central tendency – what most respondents believe
Mode	Indicates the most popular score – with the median will support data interpretation of tendency to one extreme or the other in the item
Mean	Gives an average score across the data results. Must be used with caution due to the fact that the tendency of the data is unlikely to be normally distributed
Interquartile range	Removes the extreme answers and can demonstrate if we are seeing a clear cluster or a spread of answers – a smaller number indicates clustering.

A detailed analysis was undertaken of question 1, in order to test the feasibility of the data evaluation in google sheets from the downloaded Likert item data.

The formulae were built within the spreadsheet for the four primary analyses above with the results shown in table 12.

Table 12: The constructed test data analyses for feasibility

Statistic	Formula	Result
Median	=SUM(MEDIAN(FIRSTCELL:LASTCELL))	7
Mode	=SUM(FIRSTCELL:LASTCELL)	7
Mean	=AVERAGE(FIRSTCELL:LASTCELL)	6.22
Interquartile range	*Quartile 1* =QUARTILE(FIRSTCELL:LASTCELL,1) *Quartile 3* =QUARTILE(FIRSTCELL:LASTCELL,3) *IQR* =SUM(Q1-Q3)	6 8 2

The results from the analysis of question 1 were checked by hand and confirmed as correct in order to have confidence that the analysis approach was robust and scalable to large participant numbers.

The data was then reviewed graphically to determine whether the data appeared parametric or non-parametric in distribution. This is shown in figure 7 which demonstrates a non-parametric distribution of the data.

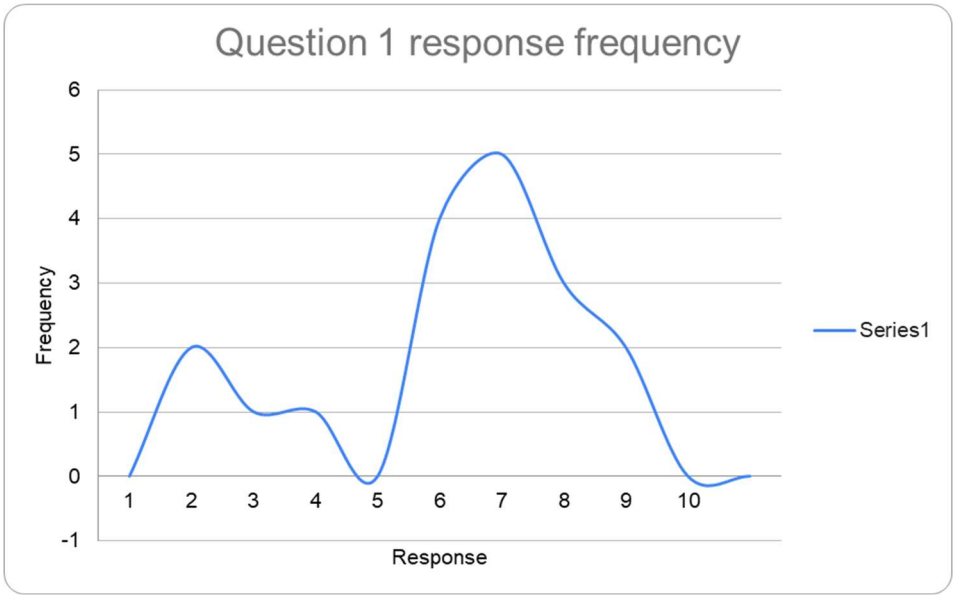


Figure 7: Data distribution graph for Question one

The nature of the data distribution is important because when the data is to be analysed comparing specific groups – particularly between types of clinician (Medical student vs Physician Associate student) and also different universities - comparative statistical analysis will need to be utilised (Rumsey, 2003). It must also be noted that the pilot analysis was on a very small data set and thus parametric analysis might not be considered feasible (Rumsey, 2003). Given the large data set expected in the main study dataset, and in line with the assumptions previously described regarding central limit theorem, a two-tailed t-test will be used for data analysis with boot strapping for sensitivity analysis as described on page 76.

In order to test the use of the two tailed t-test a comparison was done on the two sets of data from the pilot groups (groups A and B). These calculations were run in SPSS to ensure calculation accuracy. The results of the initial analysis showed a two tailed t test result p = 0.214. Given the very low numbers this served as a feasibility test of the statistical approach, rather than permitting any conclusion. The use of SPSS showed that it is straightforward to scale this up for the data from the main study and add relevant additional information for detailed analysis, including type of assessment, university, and score. These approaches will

be utilised to evaluate for differences in thinking between groups within the quantitative arm. Bootstrapping will be added as a sensitivity analysis.

5.6.3 Baseline Data – Qualitative component

As described within section 5.5.2 the pilot transcript was analysed by me against the initial coding frame, developed in the NGT process. This yielded data for the appropriateness of the initial coding framework.

Some codes were not used at all and considered to be insufficiently specific, similarly other codes were generated or split to give a more accurate representation of the content. From this a proposed series of dimensions and some initial subcategories were produced. The primary changes were the addition of a dimension for assessment centred context and the initial generation of some clear subcategories, specifically clinical setting and demographics under patient centred context and the shifting of red flags to sit as a subcategory under satisficing as, when mentioned, red flags were used as a way to agree a decision was "good enough".

The content was then reanalysed using the newly defined coding framework to assess its utility and effectiveness. Two further sub codes emerged under assessment centred context to allow the capture of information specific to either OSCE or single best answer questions. The final framework is shown in table 13.

Table 13: The final Dimensions and subcategories to go forward to the main project

Dimensions	Subcategories
Abductive reasoning	
Biases	
Deductive Reasoning	
Early Hypothesis Generation	
Assessment Centred Context	OSCE Specific
	SBA Specific
Pattern Recognition	
Metacognition	
Patient Centred Context	Clinical setting
	Demographics
Satisficing	Red Flags
Student Centred Context	

These codes were reviewed against the developed cognitive framework (figure 8) to ensure they represented questioning across its full range.

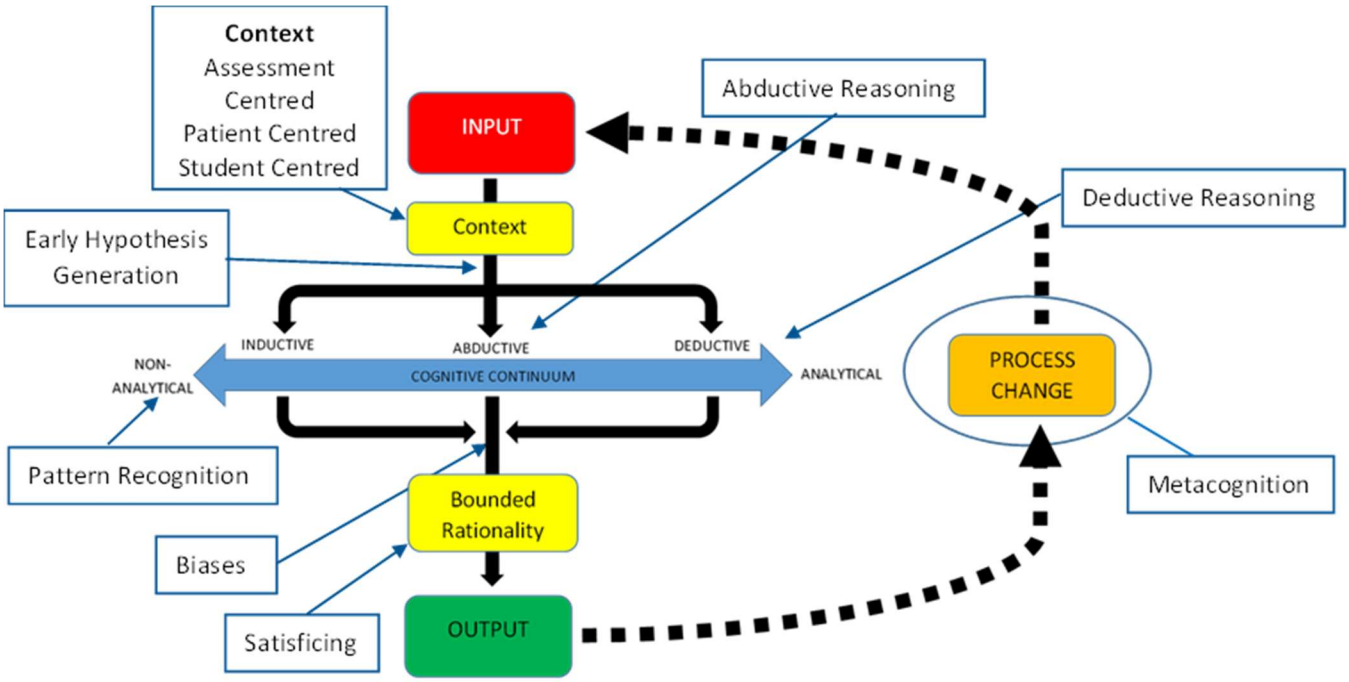


Figure 8 – Final Dimensions against the developed cognitive framework

The review demonstrated that the dimensions effectively covered the entire process and thus were fit for purpose. In this figure, "metacognition" has been linked to process change whilst recognising that it is a relatively all-encompassing dimension linked to control of thinking. The decision to place it there is due to the analysis of the pilot focus group suggesting that the concept of "process change" is linked to "regulation of thought" when the initial process seems to yield an unexpected result and thus the student reviews and repeats the cycle.

5.7 Discussion

As described in the opening of this chapter the focus of this pilot study was on feasibility at all stages of the process, from recruitment through to data capture and analysis.

Within the quantitative component of the pilot study the participant numbers (power) would never be sufficient to enable accurate statistical evaluation of responses, nor comparison of two individual groups however they permitted feasibility testing of the analysis process. In the qualitative component a single focus group could not be said to be representative of the whole, nor provide data saturation to enable meaningful conclusion to be drawn, but did enable development of the initial dimensions and some subcategories for

the analysis of further data. The feasibility is therefore considered below under the headings set out in table 9.

5.7.1 Process and Resources

The recruitment appeared fit for purpose, recognising that this was an easily accessible group to recruit. Consideration needs to be given for potential barriers when engaging students from other institutions. The pilot group participants found the SBA questions to be at the appropriate standard and the presentation of a detailed breakdown of the answers was a particular draw for participants. Consequently, it was not expected that recruitment would prove problematic in the quantitative arm. Whilst no such incentive existed for the qualitative arm it was anticipated that recruitment would be possible without difficulty, however this would need to be reviewed in the event of low number uptake. Whilst the pilot group represented only part of the target study group (physician associate but not medical students), dialogue with the local medical student group suggested that this would be welcomed, so this was not expected to be an issue.

Both data capture approaches appeared fit for purpose and were able to collect the information appropriately for further analysis. The data could be exported effectively into appropriate analysis tools for both arms of the study.

Finally, the time requirement for data capture appeared acceptable to the target group. The quantitative arm was seen as adding added value to their own revision which means it was seen as time worth investing.

Overall the feasibility of resources appeared to be good with no clear areas for concern.

5.7.2 Management

The ability to ensure that all questions are mandatory prevents data loss and helps ensure validity. A small number of errors were noted within the capture tool – all around ensuring the correct answer was identified for a question. These were easily altered and managed.

The recording media in the qualitative arm worked effectively, with the automatic transcription adding value to the process in providing the researcher with an initial framework to edit, reducing transcription time but still allowing full data immersion. The

equipment utilised was fit for purpose and should ensure risk of data loss is kept to an absolute minimum.

The transcripts were reviewed specifically regarding my input with a view to ensuring that the group was facilitated and not led. It was felt that this was the case when my inputs were reviewed via the transcripts, however this would be monitored during the transcription process of each focus group.

As space requirements would be dependent upon the group undertaking the project at any time, meaningful conclusion regarding this issue could not be reached. It was expected that all sites engaged with would have suitable spaces for the study to be undertaken.

5.7.3 Scientific

Whilst not at a level to permit the drawing of conclusions, the process enabled testing for feasibility, regarding the use of specific analytical tools for the analysis. The format facilitated the data being transferred to the relevant analysis tools and the analysis itself could be performed effectively with the packages chosen.
Both data sets appeared appropriate for the proposed methodological approach and the data should be able to be combined through the use of the same underpinning questions in both study arms.

5.8 Summary and next steps

The resources were developed using standards considered appropriate for the relevant formats including appropriate peer review to provide quality assurance that they are at a standard expected for the subject group.

Sample size calculation was performed using both parametric and non-parametric approaches and provided an appropriate minimal sample size for the testing of the statistical analyses of the quantitative component to ensure validity of the results generated.

This pilot enabled a conclusion to be drawn that the study was feasible in its proposed format and with the tools proposed for its delivery. It also supported scalability, with the current tools for both data capture and analysis, as well as appropriateness for the target group. The next step in the process was the recruitment of additional sites and groups in

order to meet the required sample sizes and target student backgrounds to enable both overall and specific group conclusions to be reached.

Chapter 6 – Research Methods

6.1 Introduction and Chapter Summary

This chapter sets out the detailed methods undertaken to capture the research data at the three sites. These methods are underpinned by the methodology set out in chapter 3 and developed from the pilot as described in chapter 5. The questions used were developed via the NGT approach which is discussed in more detail in chapter 4.

This chapter initially describes the sites used within the study, including the overall number of students from each site who participated. The next section of the chapter describes the process undertaken for the primary data collection at the different sites, initially describing the sites themselves (although the exact details of the sites has been removed for the purpose of anonymity) and the data sharing and ethics agreement process. Subsequently the chapter will consider detail of the data capture at each site for both the quantitative and qualitative elements of the study, as these had some minor differences. The chapter will conclude with a discussion regarding the practicalities of the data analysis.

Differences between sites are also discussed further under the strengths and limitations chapter (11).

6.2 Site details and Recruitment

This study relied on the participation of final year medical, and physician associate students. In order to improve validity, data collection was undertaken from three different UK sites, referred to as University 1, University 2 and University 3. This recruitment was via colleagues with a similar interest in the purpose of this study who were able to support the research through assisting with access to the student body, following agreement with them.

Final year students were selected as, whilst all medical and physician associate courses work to centrally set learner outcomes from the General Medical Council, they differ in the way that they deliver these from programme to programme. Focussing on final year students was designed to ensure that in all cases the overall knowledge and clinical reasoning development should be at a similar level to their peers as they approach the point of graduation facilitating generalisability of the work.

6.2.1 University 1

University 1 is a Russell Group University with the medical school founded in 1828 and it currently runs a five-year undergraduate and four-year postgraduate medicine course. It has also run a two-year Physician Associate course since September 2016. For the study year in question there were 240 medical students and 22 physician associate students in the year group cohorts invited to participate.

The medicine course is an undergraduate, integrated systems-based course taken across five years. The first two years are spent predominantly within the University itself with the later three years spent largely on placement. There is also a graduate entry medicine course of four years in length which is common to the five-year course from its second year. The physician associate course is a postgraduate course taken across two years. The first year is spent predominantly within the University learning the core knowledge, including lecture teaching alongside second year medical students, with a second year spent predominantly on placement in which there is contact with medical students, and some joint learning, between the two groups.

6.2.2 University 2

University 2 developed as a school of art (est. 1858) before combining with a number of other tertiary education institutions becoming a university in 1992. It has run a medicine course since 2018 and a physician associate course since 2015. For the study year in question there were 100 medical students and 48 physician associate students in the year group cohorts invited to participate. The courses both run from the same campus.

University 2's medicine course is a five-year undergraduate programme which is integrated, and systems based, similar to University 1. Students have both primary care and acute trust placements within the first two years, alongside teaching, with then further placements in the subsequent years. The physician associate course is a two-year course postgraduate course with students spending the majority of year one in the University with some GP placements. The second year is spent mainly on placement in both primary and secondary care. There is some combined teaching between the physician associate and medical students with joint simulated GP surgeries and occasional interprofessional learning events.

6.2.3 University 3

University 3 is a well-established teaching hospital institution with a medical school formally established in 1834. The physician associate course was one of the first to be set up in the UK being established in 2008 and, similar to the other two institutions, is a two-year postgraduate master's course. At University 3 it was only the physician associate course that was included in order to improve the overall physician associate numbers in the study and allow for meaningful comparison between medical and physician associate students in the quantitative arm of the study. There were 62 physician associate students in the year group cohort invited to participate.

The physician associate course at university 3 is similar to that at university 2 with students spending the first year mainly in the university with a day weekly in local GP practises. In year two they spend the majority of their time in clinical placement. There is no teaching that is run jointly with the medical course.

Table 14 shows the numbers of students in each of the student cohorts who were invited to participate in the study. Recruitment was entirely voluntary, and it was made clear to all participants that they could withdraw from the study at any time. No students dropped out after agreeing to take part in the study.

Table 14: Number of students in the invited student cohorts by site

Site	Medical Students	PA Students
University 1	240	22
University 2	100	48
University 3	[redacted]	62
Total potential participants	**340**	**132**

6.3 Ethics and data sharing considerations

The use of multiple sites required engagement with the relevant teams at each individual university to confirm they were happy with the ethics requirements and confirm the necessary agreements with regards to data sharing.

6.3.1 Ethical Approval

The Ethical approval for the main study was sought via the University 1 Ethics board and granted as per appendix 9. With the research being undertaken in the other Universities it was necessary to confirm that the ethics was deemed satisfactory to permit the study to be undertaken.

In both cases the ethics approval and all study details were provided to the participating Universities. University 2 confirmed acceptability through their Associate Professor of medical education, research, and development. University 3 confirmed through their in-house ethics team that the work could proceed. In both cases it was recognised that participation was entirely voluntary on behalf of the participants and no further ethical submission was required. Participant information forms were provided to the institutions and all potential participants, including details on who to contact in the event of any concerns. Informed consent was obtained from all participants, and it was ensured that all those invited to participate recognised that to do so was entirely voluntary.

6.3.2 Data Sharing

It was necessary to establish whether specific data sharing agreements were required between the Universities. Initial advice was sought from the University 1 medical school research team and the information governance team.

The advice provided was as follows:

My understanding is that providing the data collected as part of the research will not actually be shared with the external organisations and will only be collected via the programmes/apps/IT accounts provided by the University 1 (so they are under the control of the University) then you wouldn't need a data sharing agreement from the Uni 1 perspective. However, the external organisations may require this if data about their students is being shared with you/Uni 1. (Personal email correspondence 26/2/23).

This email can be found included as appendix 10. This guidance was provided to the other study sites, and both agreed that they were happy to proceed on the basis set out above. Consequently no data sharing agreements were required by any participating party.

6.4 Data capture practicalities – Quantitative

The process of SBA data capture was highly standardised due to the nature of this type of assessment and therefore could be run on an identical basis in each of the participating institutions. The OSCE data capture varied from institution to institution due to differing OSCE delivery approaches and set ups.

Data capture at university 2 involved a day at the university with medical students undertaking the SBA component in the morning and OSCE in the afternoon, whilst the physician associate students did the opposite. Data capture at university 1 involved four evening sessions, three dedicated to OSCE and one to the SBA for medical students, and at the formative OSCE and an afternoon SBA for the physician associate students. At University 3 the SBA data capture was undertaken in an evening session, with OSCE data captured after the students summative OSCE due to constraints on time to run it within a formative setting.

6.4.1 SBA Data Capture

The SBA data capture was conducted utilising the google form process that had been piloted previously. Individuals used an electronic device of their choice (computer, tablet, or phone) to run the google form and answer the questions. The form itself contained the participant information and consent form as well as the study questions at the end of the SBA test itself.

With the changed context in question nine of the SBA requiring a split in the group, a simple randomisation approach was taken, whereby on entering the room the students blindly selected a red or blue token and then undertook the question set (A or B) corresponding to that colour. The question set could be easily selected by scanning a QR code or entering a short URL corresponding to that question set as demonstrated in figure 9 below.

Each individual institution had its own duplicate question sets (individual Google forms) so that the baseline data could be kept separate. The number of tokens available for selection was set to correspond to the expected number at each SBA session (half red and half blue) however due to a few individuals not attending for the sessions there were a small number of tokens left over leading to a slight skew in the results numbers as described in the quantitative results.

In all institutions the questions were answered sat in an appropriate lecture theatre, or other suitable room, under exam conditions with no talking permitted to ensure that they were undertaken in as close to summative conditions as possible.

Figure 9: Example of the QR code/Short URL access point slide for students

On completion of the SBA sessions the specific question sets for that institution were closed for further responses. This was designed to prevent students choosing to “have another go” in their own time with the potential to affect the data as these would not be being carried out under the same conditions.

University 2 was the first site for delivery of the study components. During the morning SBA session a small number of issues were found on the Google form, specifically that one question set skipped over the questions as to whether they were a physician associate, or medical, student and the second was that for a couple of questions the answers had not been correctly flagged, so no correct answer was provided to the student. The first was easy to amend for the afternoon and, in addition, the google form time-stamped each entry allowing identification of the nature of the individual completing it and thus data integrity could be maintained. The question issue was amended contemporaneously, and by refreshing the form, the students got their full set of answers. No further issues were noted during the delivery of the SBAs at either University 1 or University 3.

6.4.2 OSCE Data Capture

Due to the nature of the OSCE component, this was delivered slightly differently in each of the institutions. As some sites utilised rest stations and others did not, it was decided that the use of participant information and consent forms built into the Google form would be impractical in this case. Therefore consent was collected using paper forms for the OSCE component of the study. At University 1 and University 2 iPads were utilised from the University 1 Medical School and students completed the form on these. This made the process simpler for the students and improved data capture as they were not required to log on to their own devices which might also lead to some choosing not to do so and reducing the data available.

6.4.3 University 2

The University 2 OSCE circuit consisted of four stations including the two study stations and two further, all provided by myself. The stations themselves were run in individual closed rooms or open areas with curtain separations. Because of the short circuits participants were handed an iPad after completing the OSCE circuit to undertake the quantitative research questions thereby ensuring maximum immersion in the OSCE prior to data capture.

6.4.4 University 1

The University 1 Medical Students OSCE circuit, consisted of twelve stations, including the two study stations, and two rest stations. The rest stations were at stations 6 and 12 with the study stations at stations 5 and 11. The OSCEs were run in three large open lecture theatres with the circuits spread around the available space. IPads with the questions on were located on the rest stations and students were asked to complete the data capture at their second rest station. This ensured that every individual completing the form had done at least one of the study stations (only someone starting on a rest station would not) and all had done a minimum of six OSCE stations in total.

The physician associates undertook a formative OSCE under full exam conditions with the stations arranged on the same basis as above. IPads were again placed on the rest stations for the data capture with students asked to complete the questions at their second rest station.

6.4.5 University 3

It was not possible to find an opportunity to undertake the study within a formative OSCE on the physician associate course at university 3. Following discussion with the course director it was agreed that instead, we would ask students to complete the questionnaire on completion of their summative OSCE. Whilst it was not possible to include the study stations into the summative OSCE, the exam was at the same standard as the study stations with all stations reviewed by myself as the external examiner for the programme. There was an appropriate mix of stations to drive reasoning. On completion of the assessment students returned to their holding room where they completed the questions if they were happy to participate.

6.5 Data capture practicalities – Qualitative

Focus group data acquisition was a balance of planning and practicality as outside of university 1 it relied on arranging meetings with students around the quantitative activity. Consideration was given to ensuring a balance of individuals within the groups to ensure representativeness of the groups as a whole. Field notes were taken following each group.

Unlike the pilot study, it was decided to undertake the data capture using Google Meet (Google, 2022a), which was already approved for use by University 1, easier to manage in the external environment, and run on an encrypted laptop, with no video capture, and an attached perimeter microphone. Backup recording was made using pixel record on a Pixel 7 phone. Initial concerns about the use of mobile phones for recording were explored with the IT team regarding encryption and suitability with the following advice provided:

"Pixel recorder relies on AI onboard the device, so no information is exchanged or processed in the cloud - where we would have concerns if this was the case. As long as your device is secured by password/PIN and only used for work purposes - i.e. only your university google account is in use then this workflow for interviews is actually more secure than an encrypted digital recorder used in other areas of the Faculty. I have absolutely no problem with you using the Pixel for interviews."

The use of these two options also permitted the production of a transcript which, although quite variable in its accuracy provided a useful starting point. This transcript was then edited through listening to the recording in order to ensure accuracy.

The focus groups are described below in relation to their make up in each of the individual sites. Each focus group consisted of myself and the participants only.

6.5.1 University 2

At University 2 a single focus group was run at the end of the day in which the quantitative data was captured. Because this was opportunistic it was run as a single group with volunteer participants who had undertaken both the SBA and OSCE during the day. Eighteen participants took part in this focus group (13 physician associate students and 5 medical students), above the ideal numbers in the literature, but not too many to make group management overly difficult. It did however mean that it was impossible to clarify on the recording if the speaker was a medical, or physician associate, student.

6.5.2 University 1

Final year medical students and physician associate students were contacted at the University 1 and volunteers sought. Six Physician Associate students came forward and formed a single focus group. The medical students were on placement doing student assistantships and so focus groups were run at four sites in three local hospital trusts with the local education teams, or the medical school, supporting the use of a room for the purpose. Medical students formed four focus groups of two (Site 1), three (Site 2) and four (Sites 3 and 4).

6.5.3 University 3

Physician associate students at University 3 took the SBAs in an afternoon session as part of the run up to their summative examinations. Subsequently, volunteers were sought to stay and form a focus group. Nine agreed to stay for the focus group although one member had to leave after around 40 minutes of the process.

6.6 Data analysis practicalities – Quantitative

A combination of Microsoft Excel and SPSS were utilised in order to undertake the statistical evaluation of the quantitative data.

Data collection was directly into the Google form with data then directly exported from the form into a google sheet (built in Google functionality). This data was then saved into Excel (to produce the relevant graphs and baseline statistics) and exported to SPSS to allow for effective detailed analysis (comparison and bootstrapping) against the various subsets of interest.

Comparison testing of the means of the subsets of interest was undertaken through a 2-tailed t-test with bootstrapping run as a sensitivity analysis. Graphical representation of the Likert scale data was produced in Excel with box and whisker plots (Anonymous, 2014) which permit effective visual representation of quartiles, interquartile range and overall answer spread, including specific outliers. Validity was provided using an appropriate statistical analysis of the data as previously described.

When bringing the two data sets together in the convergent analysis it was useful to consider the Likert data items narratively. For the purposes of interpretation and comparison of the quantitative data with the qualitative data the 10 point Likert data scale was considered against the typical descriptors for a five point scale as shown in table 15 (Sullivan and Artino, 2013). This facilitated description of the quantitative results for comparison with the qualitative results.

Table 15: Likert scale descriptors for interpretation

Likert Scale Score	**Descriptor term**
1 or 2	Never
3 or 4	Rarely
5 or 6	Sometimes
7 or 8	Often
9 or 10	Always

It was important to have these descriptors as, whilst we have the statistics, the scale is from "never" to "always" and so it could reasonably be asked as to what terms such as "rarely" mean in practical terms. The conversion to these terms allows a better narrative description of the quantitative results facilitating convergence with the qualitative data.

6.7 Data analysis practicalities – Qualitative

Qualitative data was analysed using NVIVO release 1.7.1. The initial transcripts were converted to Microsoft Word files and then edited through a process of immersion and text editing until an accurate final transcript for each focus group was produced.

These transcripts were analysed by myself utilising the initial coding framework produced at the pilot stage with the number of new subcategories generated after each analysis recorded as a proxy measure for saturation. Details of this coding process is included within chapter 8 - the qualitative results.

Validity of the qualitative results was ensured by revisiting the key measures of validity set out in table 2 within chapter 3, specifically:

- Descriptive validity – the research will be factually accurate through the process of transcription and coding process.
- Interpretive validity – The data will be analysed to capture the true meaning and understanding of the participants and laid out clearly in relation to the conceptual framework.
- Theoretical validity – The use of the concept framework to set out the results will enable the research to be linked to the conceptual framework and thus explain the phenomena under investigation.
- Transparency – The use of thematic analysis, and working through its stages, will ensure transparency in the process.

Generalisability (that the research will be useful to understand other situations) and evaluative validity (that a judgement is made on the subject) will be part of the qualitative results, set out in chapter 8, and also be a core part of the mixed methods convergent analysis and interpretation contained within chapter 9.

6.8 Summary

The methods used for the data capture and interpretation followed from the methodology with a combination of planning and pragmatic opportunity facilitating the different elements of the study at the differing sites. The data capture was successful, and data managed as described above. Detailed quantitative and qualitative data management is

discussed in the relevant chapters 7 and 8. Limitations and mitigations within the methods are discussed in detail in chapter 11.

Chapter 7 – Quantitative Results

7.1 Introduction and Chapter Summary

This chapter sets out the results from the quantitative data collection undertaken at the three study sites. The demographics of the sample populations will first be discussed with breakdown by site, assessment type, clinical student group, gender and ethnicity. This will set out the data obtained in this part of the study. The primary results will then be discussed with respect to the data as a whole, broken down by the two different assessment types with a comparison made between them and summaries of the results provided.

The data will then be broken down to undertake comparison between subgroups, specifically clinical student type, gender and ethnicity, with calculations made for each of the two assessment types.

The statistical analyses in this chapter are conducted in line with the methodology and method as set out in chapters 3 and 6. This chapter will include interpretation of these results with chapter 9 seeking to bring together the quantitative and qualitative results in an interpretive discussion, as part of a convergent process, and draw inferences from them.

The sample for analysis consists of Likert data items collected from the SBA and OSCE assessments undertaken by students at the three schools involved as set out in chapter 6. For the purposes of this chapter a data point is a complete set of 10 questions answered by an individual student for either the SBA or OSCE tests.

Finally the chapter will review the findings of the overall results and comparative groups, concluding with a short summary.

For clarity, figure 8 shows how each question relates to the originally developed concept framework.

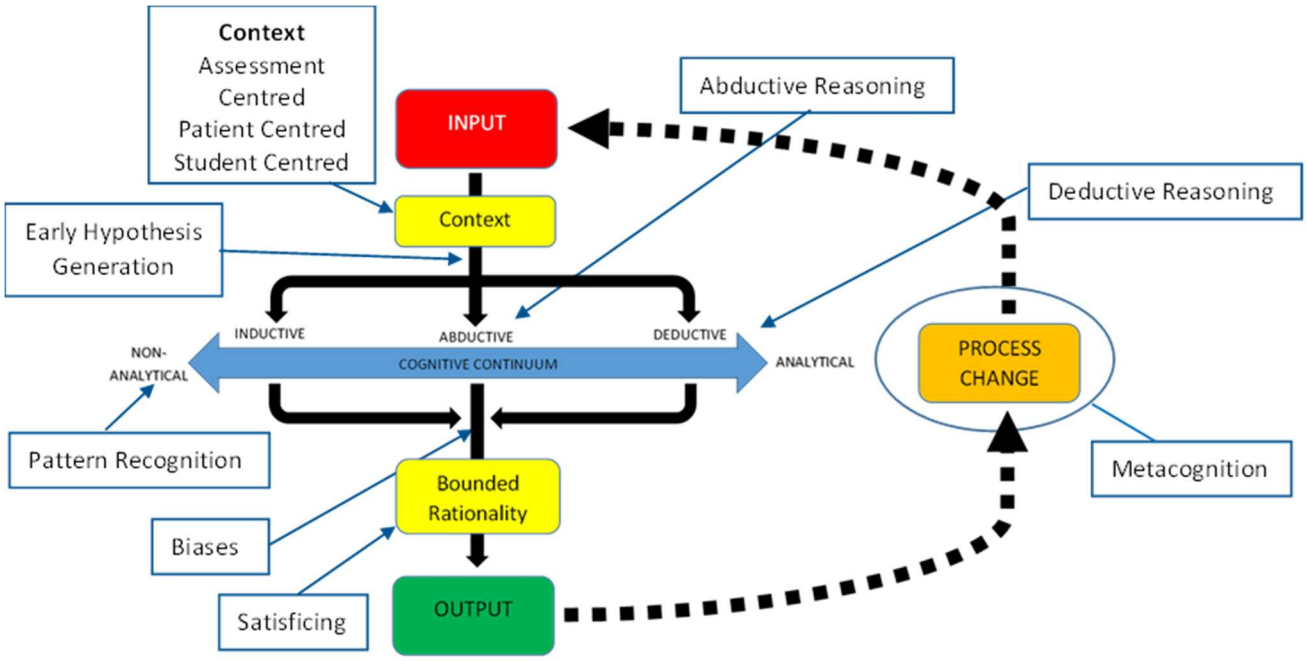


Figure 8 – Final Dimensions against the developed cognitive framework

7.2 Demographics

The three different sites produced a total of 595 separate data points for SBA (Single Best Answer) and OSCE (Objective Structured Clinical Examination). Given that the majority of the participants will have completed both the SBA and OSCE, the data is expressed in terms of data points. The breakdown of the overall data points (full sets of 10 question responses) by site and examination type are set out in table 16.

Table 16: Overall data points by study site and examination type

Site	Exam Type	Total Data Points
University 1	SBA	174
	OSCE	216
University 2	SBA	63
	OSCE	58
University 3	SBA	54
	OSCE	30
	Total	**595**

This data gives overall numbers by exam type of 304 responses pertaining to OSCE and 291 pertaining to SBA and represents the entire study quantitative data. These overall numbers are slightly below the ideal numbers based on the sample size calculation however still represent a good sample size for each of the two examination types.

7.2.1 Clinical Role

Data was captured on the study participants clinical role (medical student or student physician associate). This data is set out in table 17.

Table 17: Number of data points by clinical role

Clinical Role	Total Data Points	Percentage of Responses
Medical Student	400	67.2
Student Physician Associate	195	32.8

It should be noted that these proportions have higher physician associate representation than would be expected from the numbers in both training and employment for the two professions. This was to permit meaningful comparison between the two groups in more detailed analysis.

7.2.2 Examination type

The two examinations used were Single Best Answer and OSCE. Overall, slightly higher numbers participated in the OSCE testing opportunity. The number of data points for each assessment type are shown in table 18.

Table 18: Number of data points by examination type

Examination	Total Data Points	Percentage of Responses
SBA	291	48.9
OSCE	304	51.1

Overall uptake was good, and the figures are relatively similar between the two groups. To obtain meaningful comparison between medical students and student physician associates it is necessary to have sufficient data points for the two groups. This data is shown in Table 19.

This data shows that whilst student Physician Associate numbers are lower, they are all at a level to permit a reasonable comparison to be made based upon the previously described methodology.

Table 19: Number of data points by examination type and role

Role	Examination	Total Data Points	Percentage of Responses
Medical Student	SBA	181	30.4
	OSCE	219	36.8
Student Physician Associate	SBA	110	18.5
	OSCE	85	14.3

7.2.3 Gender

Of the 595 responses three of the responses (0.5%) stated that the gender that they associated with was not the one assigned at birth. Table 20 shows the gender split of the data points.

Table 20: Number of data points by gender

Gender	Total Data Points	Percentage of Responses
Male	168	28.2
Female	422	70.9
Prefer not to say	5	0.8

It is notable that the gender balance in the results falls in line with statistics for entry for medicine with UCAS (University and Colleges Admission System) figures for 2022 Medicine entry showing offers made to 13,580 females vs 6,815 males (66.6% vs 33.4%)(UCAS, 2022). Physician Associate studies, whilst a relatively new profession in the UK, has been female dominated with a 2022 study showing female to male split of 85.1% vs 14.9% (Roberts *et al.*, 2022). The breakdown of gender by role for the data points in this study is shown in table 21. This demonstrates that the percentage splits within the genders was as expected based upon the known data regarding entry to these clinical professions.

Table 21: Overall data by gender and role

Role	Gender	Number	Percentage for profession	Percentage Overall
Medical Student	Male	138	34.5	23.2
	Female	257	64.3	43.2
	Prefer not to say	5	1.3	0.8
Student Physician Associate	Male	30	15.4	5.0
	Female	165	84.6	27.7

No individual could complete the research questions more than once for each of SBA and OSCE therefore as a sense check on the gender proportions we can also look at the gender data for each of the two exam types as set out in table 22.

Table 22: Overall data by gender and examination type

Examination	Gender	Number	Percentage for Exam type	Percentage Overall
SBA	Male	74	25.4	12.4
	Female	212	72.9	35.6
	Prefer not to say	5	1.7	0.8
OSCE	Male	94	30.9	15.8
	Female	210	69.1	35.3

These also approximate to the gender proportions expected from what is known regarding the make-up of the professions.

7.2.4 Ethnicity

Ethnicity data was captured based upon the UK government census groups as previously discussed. This is set out in Table 23.

The UK government considers all those who do not fall within the “White - Scottish/English/Welsh/Northern Irish/British” group to be ethnic minorities. BAME, or BME, is defined as all ethnicities other than white groups. It is important to note that this is the terminology used within groupings in other literature pertaining to exam pass rate variability between white and BAME groups. On this basis it was decided that this would be the way the data would be managed for the purposes of analysis to enable the comparison

between groups of a meaningful size (Rimmer, 2016). From this point forwards it will be these combined ethnicity groups of "White" or "BAME" that will be used for analysis.

Table 23: Overall data points by ethnicity

Ethnicity	Data Points	Percentage of Responses
White - Scottish/English/Welsh/Northern Irish/British	303	50.9
Asian/Asian British - Indian	46	7.7
Black/African/Caribbean/Black British - African	44	7.4
Asian/Asian British - Pakistani	38	6.4
Asian/Asian British - Any other	35	5.9
Asian/Asian British - Bangladeshi	28	4.7
Other - Any other ethnic group	19	3.2
White - Any other	17	2.9
Mixed Multiple - White and Asian	12	2.0
Asian/Asian British - Chinese	11	1.8
Prefer Not To Say	9	1.5
Mixed/Multiple - Any other	9	1.5
Other - Arab	8	1.3
White - Irish	5	0.8
Black/African/Caribbean/Black British - Caribbean	5	0.8
Mixed/Multiple - White and Black Caribbean	3	0.5
Black/African/Caribbean/Black British - Any other	3	0.5

The data points based upon this division are shown in table 24.

Table 24: Overall data points by large ethnicity group

Ethnicity Group	Data Points	Percentage of Responses
White	325	54.6
BAME	261	43.9
Prefer Not To Say	9	1.5

It is important to understand the data in the same terms as the gender data, specifically regarding the difference by profession and by exam type. This breakdown is shown in tables 25 and 26.

It is notable from these results that the physician associate group has much greater BAME representation compared to the medical group, indeed it is almost a mirror image. This may represent the fact that physician associate studies is predominantly a postgraduate profession and thus those may who have struggled at A-level due to educational

disadvantage (Sammons *et al.*, 2015) have been able achieve success in their studies at University and therefore can pursue this as a clinical career option whilst their opportunities to study medicine are much more limited.

Table 25: Overall data by ethnicity group and role

Role	Ethnicity	Number	Percentage for profession	Percentage Overall
Medical Student	White	254	63.5	42.7
	BAME	141	35.3	23.7
	Prefer not to say	5	1.3	0.8
Student Physician Associate	White	71	36.4	11.9
	BAME	120	61.5	20.2
	Prefer not to say	4	2.6	0.7

Table 26: Overall data by ethnicity group and exam type

Examination	Ethnicity	Number	Percentage for exam type	Percentage Overall
SBA	White	151	51.9	25.4
	BAME	134	46.0	22.5
	Prefer not to say	6	2.1	1.0
OSCE	White	174	57.2	29.2
	BAME	127	41.8	21.3
	Prefer not to say	3	1.0	0.5

It is also likely to be influenced by the site itself. University 2 is a new medical school specifically set up to champion widening participation which particularly encourages BAME applicants. Physician associate courses recruit heavily from the local area and University 3 is within an area of relatively high ethnicity (46% white, 6% mixed, 29% Asian, and 18% black) compared to the UK general population. To explore in more depth the full data set was analysed by study site and course. Most students completed both assessments and so reviewing the examination type with the larger data set at each site provides the best impression of the course ethnicity mix. OSCE was used at university 1 (n=216) and SBA for University 2 (n=63) and University 3 (n=54). This is shown in table 27.

Table 27: Overall data by Site, ethnicity group and role

Site	Ethnicity	Number	Percentage for site	Percentage Overall
University 1	White	147	68.1	44.1
	BAME	66	30.6	19.8
	Prefer not to say	3	1.4	0.9
University 2	White	16	25.4	4.8
	BAME	47	74.6	14.1
University 3	White	14	25.9	4.2
	BAME	39	72.2	11.7
	Prefer not to say	1	1.9	0.3
	Total	333		

7.3 Decision-making Process Primary Results

The data for the ten questions was reviewed and the data graphically represented using box and whisker charts. For the purposes of presentation these results will be presented by examination type and by individual question.

The box extends from the 1st quartile to the 3rd quartile so 50% of the data falls within this box. The line in the box represents the median and the X in the box the mean. The whiskers extend to the maximum value and the minimum value respectively. Some points can be considered outliers if they fall outside a range greater than 1.5 times the interquartile range. These are represented as dots with the whisker extending to the highest or lowest non-outlying score.

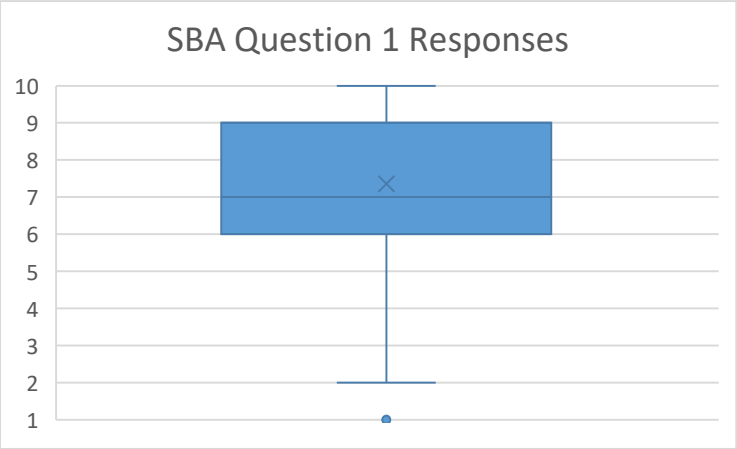
For all of the questions the selection of a score of 1 indicated a “never” response whereas a score of 10 indicated an “always” response.

The charts start at 1 rather than zero as this was the lowest possible selectable number on the Likert data items. On this basis the neutral or midpoint on the scale is 5.5. For the purpose of considering the Likert responses in narrative terms (as discussed in chapter 6) the mean will be used.

7.3.1 Single Best Answer Question responses

Question 1

Question one: How I feel in the assessment affects the way I make decisions in questions.



Median – 7

Mode – 7

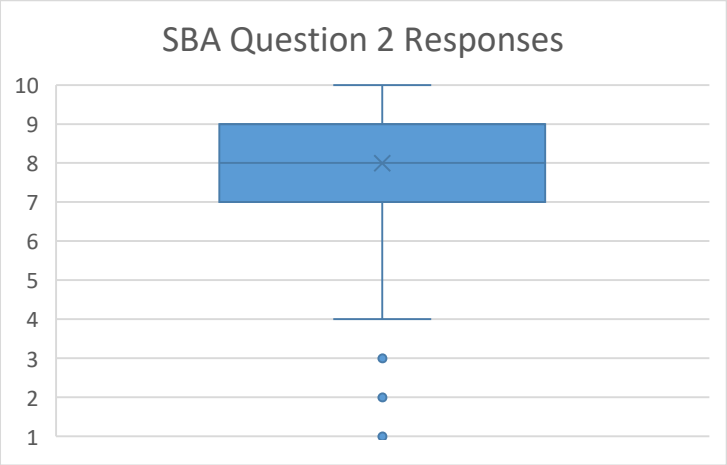
Mean – 7.36

Interquartile range - 3

The data showed that in SBA questions, how students felt often affected the way that they made decisions in this type of assessment, with over 75% of students scoring 6 or above, suggesting that the context of being in an examination itself exerts an influence on the exam process.

Question 2

Question two: The patient demographics have an impact on my decision-making (age, sex etc).



Median – 8

Mode – 10

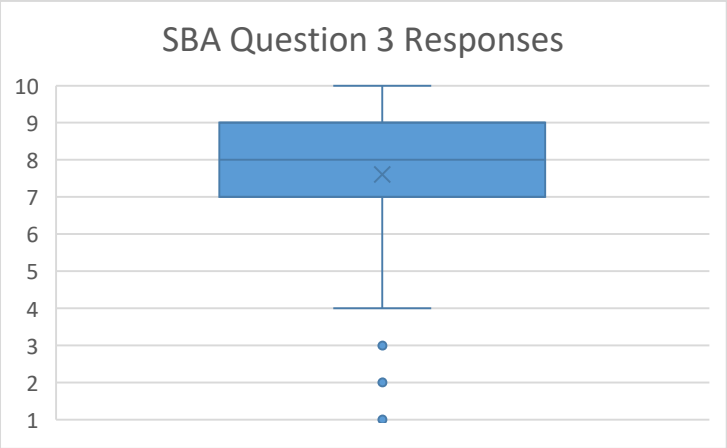
Mean – 8.00

Interquartile range - 2

The data showed that in SBA questions demographics often have a significant impact on student decision-making. The most popular score selected was 10 with a narrow interquartile range between 7 and 9. The data fits with a strong influence of these contextual factors within the question.

Question 3

Question three: The clinical setting of the question has an impact on my decision-making (GP, Emergency Department, Ward etc).



Median – 8

Mode – 8

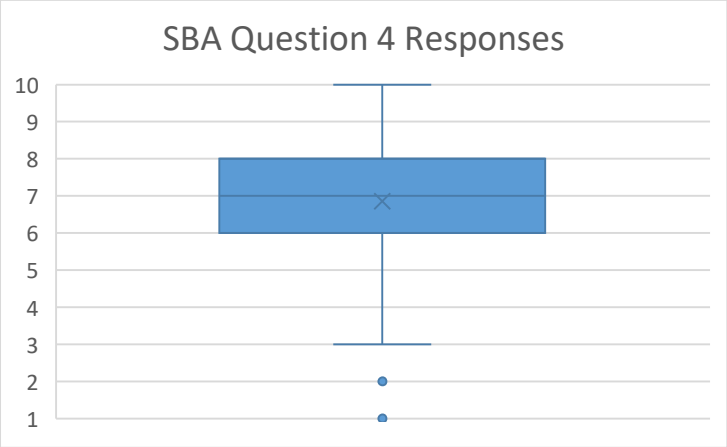
Mean – 7.6

Interquartile range - 2

The data showed that in SBA questions the setting of the question often has an impact on their decision-making however slightly less so than demographics with a mode of 8. This is again linked to contextual factors having a strong influence on the decision-making.

Question 4

Question four: I usually come up with hypotheses immediately then change them as the question develops.



Median – 7

Mode – 8

Mean – 6.86

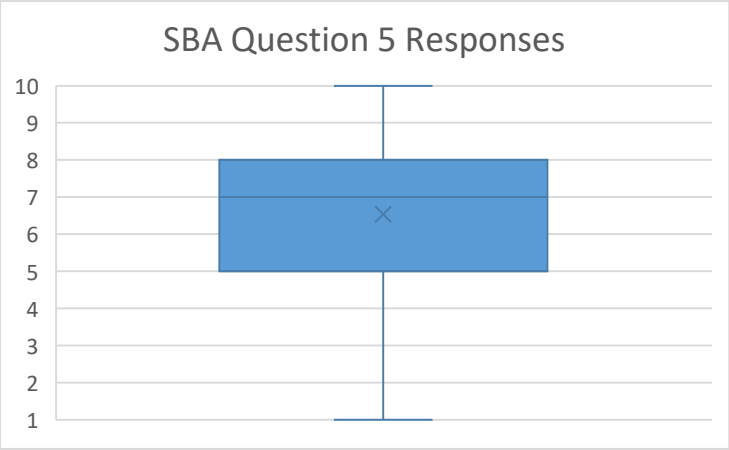
Interquartile range - 2

The data showed that in SBA questions, students sometimes tend toward early hypothesis formation with a mean of 6.86. The mode of 8 is notable as is the median of 7 especially with a relatively tight interquartile range.

Question 5

Question five: I usually generate a list of possible answers and eliminate them one by one.

This question is also linked to the concept of early hypothesis formation, but with abductive decision-making. Students indicated that they sometimes did this with a mean of 6.54. The spread of the results (with all responses within the quartiles) demonstrates that this question is one in which students show a wide range of opinion with no true outliers.



Median – 7

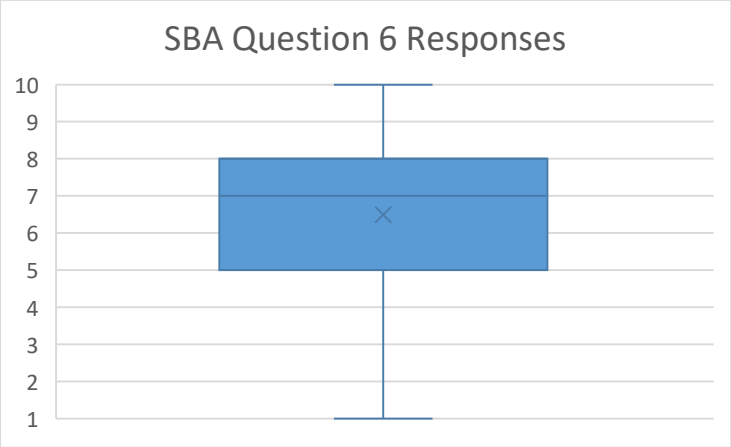
Mode – 7

Mean – 6.54

Interquartile range - 3

Question 6

Question six: I try and keep an open mind until I have formulated an answer.



Median – 7

Mode – 7

Mean – 6.49

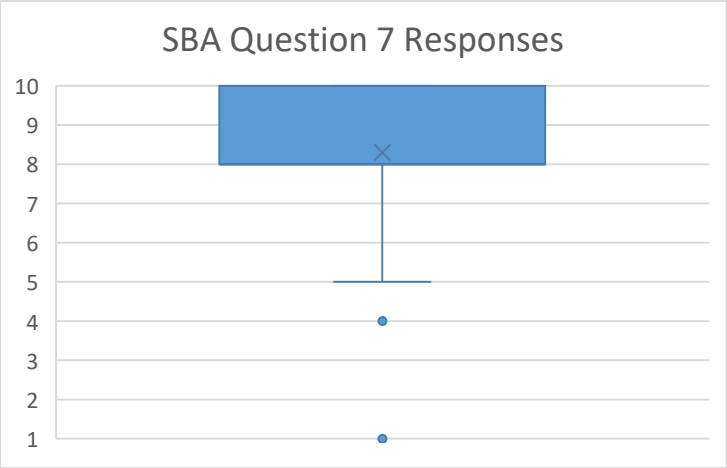
Interquartile range - 3

The data showed that in SBA questions students sometimes tend toward trying to keep an open mind until they formulate an answer. Like the previous question the spread of the results demonstrates that this question is one in which students show a wide range of opinion with no true outliers.

Question 7

Question seven: I look for recognisable patterns within the question to generate my answer.

Students have a strong tendency toward pattern recognition in SBA questions with the mean suggesting this almost always occurs. This resonates with the responses to questions 2 and 3. The narrow interquartile range and high mean allows confidence in this view and suggests a strong preference for inductive decision-making processes in this examination type.



Median – 8

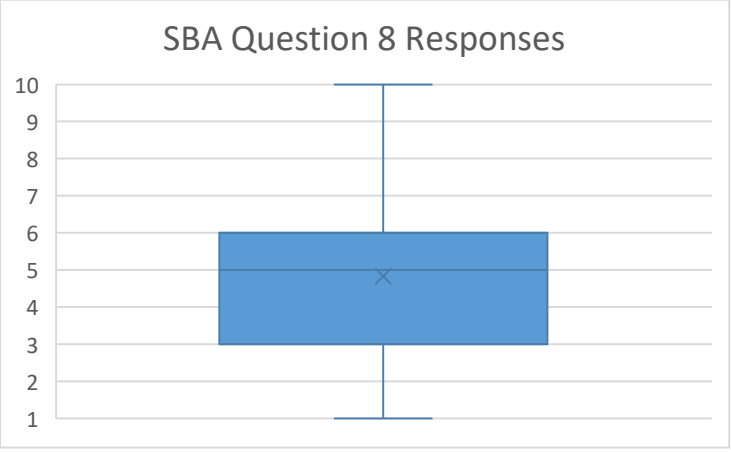
Mode – 8

Mean – 8.30

Interquartile range - 2

Question 8

Question eight: I tend to come to an answer but am not sure how I got there.



Median – 5

Mode – 4

Mean – 4.83

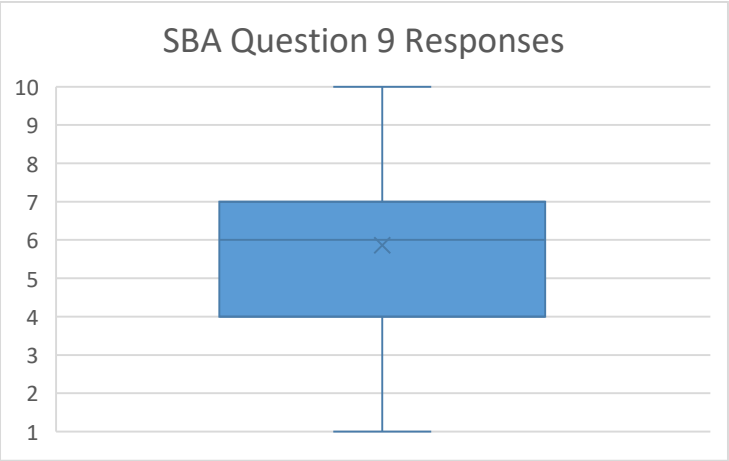
Interquartile range - 3

Students tended toward feeling that they were clear on how they came to an answer in the majority of cases with the data indicating students rarely felt that they tended to come to an answer but that they were unclear how. The wide quartile spread is notable with a small number of individuals finding this happened a lot. The data suggests that for most there was

a clear process which might be either pattern recognition or deduction as per previous answers.

Question 9

Question nine: I aim for a "good-enough" answer rather than a perfect one.



Median – 6

Mode – 7

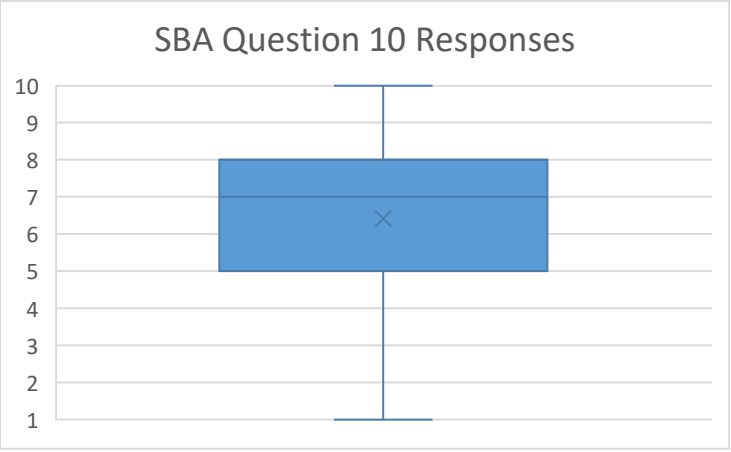
Mean – 5.86

Interquartile range – 2.5

The data suggest students sometimes aim for good-enough with the mean sitting near the midpoint of the scale. It is notable that the mode of 7 suggests that there is some tendency toward a bounded rationality approach in the answering of questions.

Question 10

Question ten: I acknowledge and try to avoid cognitive biases in my thinking.



Median – 7

Mode – 7

Mean – 6.42

Interquartile range - 3

The responses show students often acknowledge and seek to control biases in an attempt to try and avoid them becoming decision affecting. The wider interquartile range (which crosses the midpoint) and wide spread of answers suggests a wide variance in students' perception of this issue.

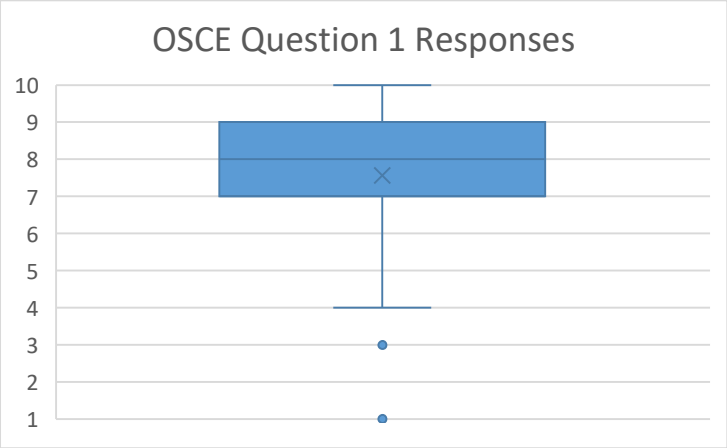
Summary of Findings in SBA responses

Relating the findings back to the concept framework, the quantitative results suggest that students use contextual factors such as demographics and setting as part of a pattern recognition process as their primary approach to decision-making in assessments. The data also demonstrates that students feel that they are affected in their thinking processes simply by being in an examination situation. The strong tendency toward pattern recognition fits with early hypothesis formation however, despite this, there is evidence of both deductive and abductive decision-making, albeit less strongly. Bounded rationality is evidenced, however relatively weakly and students have awareness of the potential impact of biases in their answers.

7.3.2 OSCE Question responses

Question 1

Question one: How I feel in the assessment affects the way I make decisions in questions.



Median – 8

Mode – 8

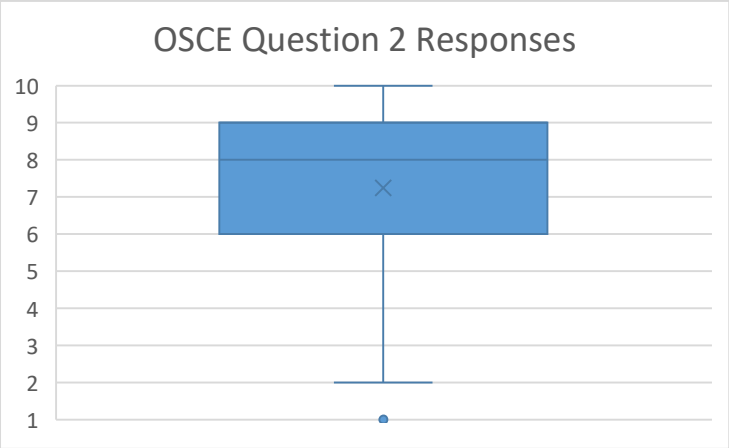
Mean – 7.59

Interquartile range - 2

The data suggest students sometimes aim for good-enough with the mean sitting near the midpoint of the scale. It is notable that the mode of 7 suggests that there is some tendency toward a bounded rationality approach in the answering of questions.

Question 2

Question two: The patient demographics have an impact on my decision-making (age, sex etc).



Median – 8

Mode – 8

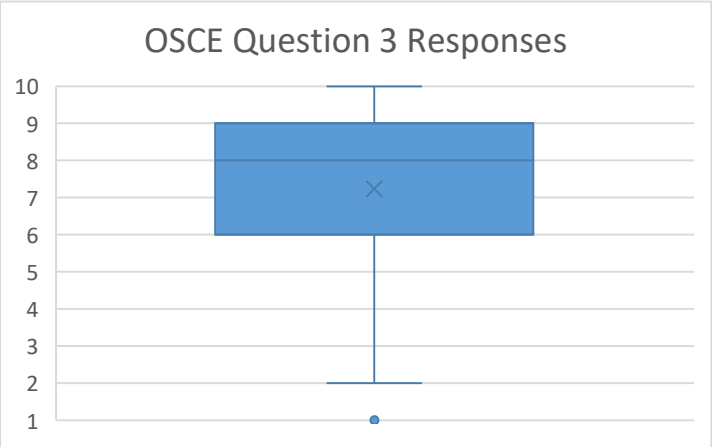
Mean – 7.26

Interquartile range - 3

In OSCE, students demonstrated a clear tendency toward impact from the demographics of the scenario, however this was less strong than for SBA questions. The overall data spread was also wider than for the SBA questions suggesting less homogeneity of view from students.

Question 3

Question three: The clinical setting of the question has an impact on my decision-making (GP, Emergency Department, Ward etc).



Median – 8

Mode – 8

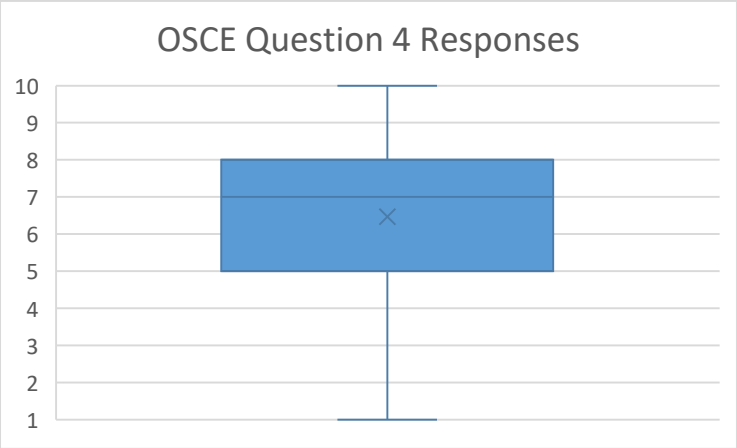
Mean – 7.27

Interquartile range - 3

The data showed that in OSCE questions, the setting of the question often impacts on their decision-making, which is identical in weighting to that of the demographics. Like question 2 the spread of responses is wider than for SBAs.

Question 4

Question four: I usually come up with hypotheses immediately then change them as the question develops.



Median – 7

Mode – 7

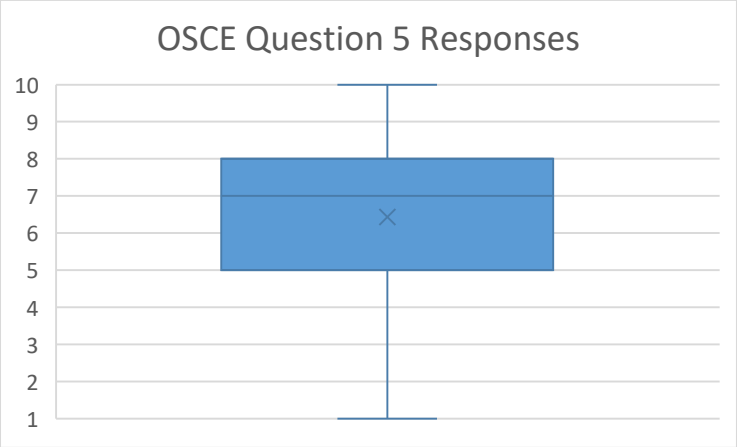
Mean – 6.52

Interquartile range - 3

The data showed that in OSCE questions, as in SBA questions, students sometimes undertake early hypothesis formation. As seen in previous questions the overall spread of the results suggests less homogeneity of student responses.

Question 5

Question five: I usually generate a list of possible answers and eliminate them one by one.



Median – 7

Mode – 8

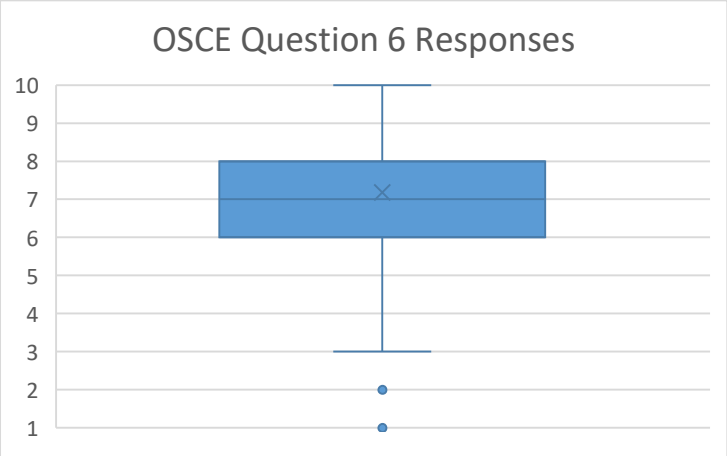
Mean – 6.49

Interquartile range - 3

The results for OSCE are almost identical to that of SBA with students sometimes taking this approach with a mean of 6.54. The spread of the results (with all responses within the quartiles) demonstrates that this question is one in which students show a wide range of opinion with no true outliers.

Question 6

Question six: I try and keep an open mind until I have formulated an answer.



Median – 7

Mode – 8

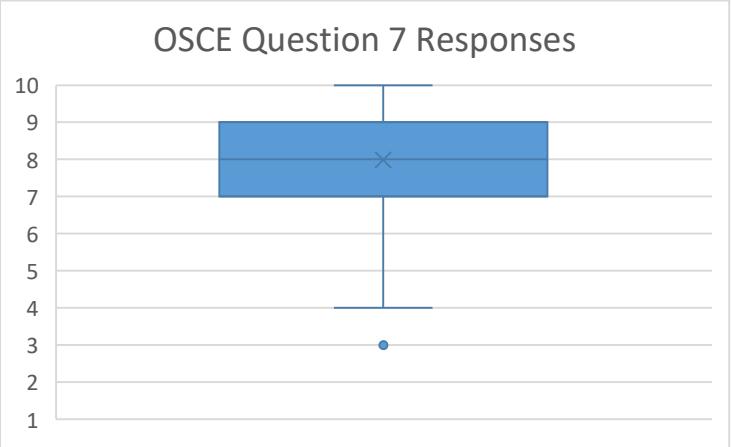
Mean – 7.18

Interquartile range - 2

When comparing the responses with those from the SBA, students show a stronger tendency toward trying to keep an open mind in OSCE examinations than within SBA questions, with the results suggesting this often occurs. Compared to SBAs the data has a narrower spread suggesting that there is a general higher tendency toward this thinking approach.

Question 7

Question seven: I look for recognisable patterns within the question to generate my answer.



Median – 8

Mode – 8

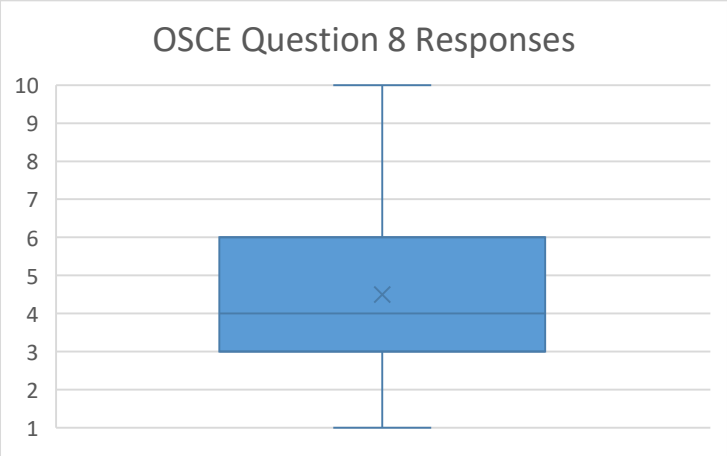
Mean – 7.98

Interquartile range - 2

Similarly to SBA data, students have a strong tendency toward pattern recognition in OSCE. However, it is noticeable that the mean sits slightly lower than in question 7 for the SBA (sitting at often, just below the always range) as does the interquartile range. Pattern recognition remains important in OSCE, but less strongly than in SBA.

Question 8

Question eight: I tend to come to an answer but am not sure how I got there.



Median – 4

Mode – 4

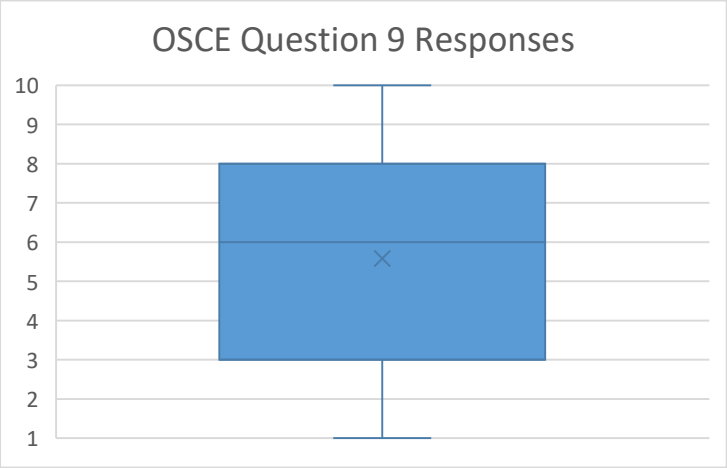
Mean – 4.53

Interquartile range - 3

Responses of students following the OSCE were similar to the responses following the SBA questions. The data indicates that students were clear on how they came to an answer in the majority of cases and only rarely feeling that they were unclear how they had got there. Like SBA there was a wide quartile spread suggesting a wide range of student opinion on this topic.

Question 9

Question nine: I aim for a "good-enough" answer rather than a perfect one.



Median – 6

Mode – 8

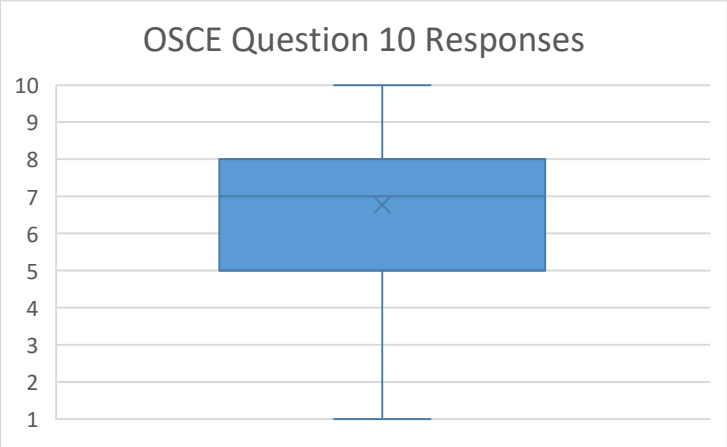
Mean – 5.57

Interquartile range – 5

The responses here showed the largest interquartile range with the mean sitting almost exactly on the midpoint of the scale with students sometimes thinking this. This suggests that students are evenly split around this issue, specifically inductive decision-making, with no specific tendency one way or the other.

Question 10

Question ten: I acknowledge and try to avoid cognitive biases in my thinking.



Median – 7

Mode – 7

Mean – 6.78

Interquartile range - 3

Like the SBA the data shows students are sometimes aware of biases and make an attempt to try and avoid them becoming decision affecting. Whilst tendency toward this is seen, the data spread is such that a suggestion that students understand biases well cannot be inferred.

Summary of Findings in OSCE responses

The results for OSCE are extremely similar to the results for SBA, although there is slightly reduced tendency toward pattern recognition which is linked to the fact that demographics appear to have a reduced influence on clinical reasoning in OSCE. The results demonstrate that students have a higher tendency to keep an open mind in OSCE assessments when compared to SBA questions, which may be linked to the more exploratory nature of the examination, with students having to discover information for themselves, compared to an SBA question where the information is directly provided. Whilst the tendency toward pattern recognition is lower in OSCE than for SBA question it remains high and, along with the influence of being in the exam itself, is still dominant in student reasoning approaches.

7.4 Comparative Analyses

Whilst the results presented above appear broadly aligned it is important that more formal statistical analysis is undertaken to compare the SBA and OSCE data results and see if any of these differences reach statistical significance.

It is also important to undertake subgroup analyses to see if any difference is observed between the different professional student groups, genders and ethnicities, which may have

an impact on potential performance in examinations. Comparison was made between the means of the question answers for the differing groups.

Analysis was with independent-t testing supported by bootstrapping for sense checking the results as previously described in chapter 5 (p76). Levene's test of variance was applied to understand the overall variance of the data sets and are included in the results (Levene, 1960). 2-sided t-test results were utilised to understand potential variance at either end of the distribution curve.

7.4.1 Overall Comparison of SBA and OSCE

Analysis was undertaken to compare the overall dataset for SBA and OSCE as per table 28.

Levene's test for variance demonstrates statistically significant variance in the distribution of responses for questions 2 and 9 only. The rest can be considered equal in variance.

The closeness of the bootstrap result to the two-sided p result gives us confidence that the use of the parametric test was appropriate, and can be considered accurate, for the purposes of drawing conclusion from the data.

Table 28: Comparison of SBA and OSCE data for difference

Question	Levene's test significance	t value	Two-sided p result	Bootstrap
1	0.825	1.308	0.191	0.185
2	<0.001	-4.281	<0.001	<0.001
3	0.32	-2.263	0.024	0.017
4	0.793	-2.464	0.014	0.019
5	0.964	-0.636	0.525	0.539
6	0.884	4.486	<0.001	<0.001
7	0.649	-2.432	0.015	0.021
8	0.687	-1.087	0.071	0.069
9	<0.001	-1.439	0.151	0.149
10	0.819	2.197	0.028	0.035

A statistically significant difference between SBA and OSCE is seen in questions 2,3,4,6,7 and 10. This fits with the observation from the analysis of the individual questions above, with the t-value results correlating with the direction of difference (positive indicates a higher OSCE result vs SBA). The large sample size allows for relatively small differences to be found, which is the case here.

We can therefore reasonably conclude from this comparison that there are statistically significant differences in the way students' reason in SBA and OSCE, however these differences are small and highlight the need to explore this in more depth through the wider statistics presented above and via the qualitative analysis.

7.4.2 Impact of clinical role in different examination types

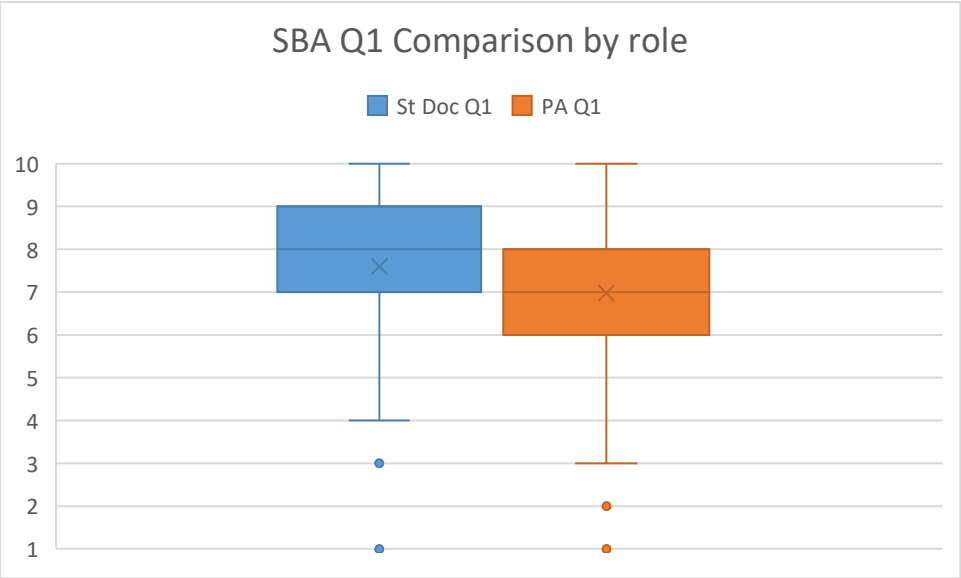
Analysis was undertaken to compare the separate datasets for SBA and OSCE. The test was looking for any statistically significant difference in the question responses between medical students and student physician associates.

For SBA there were 181 medical student data points and 110 student physician associate data points. For OSCE analysis there were 219 medical student data points and 85 student physician associate data points.

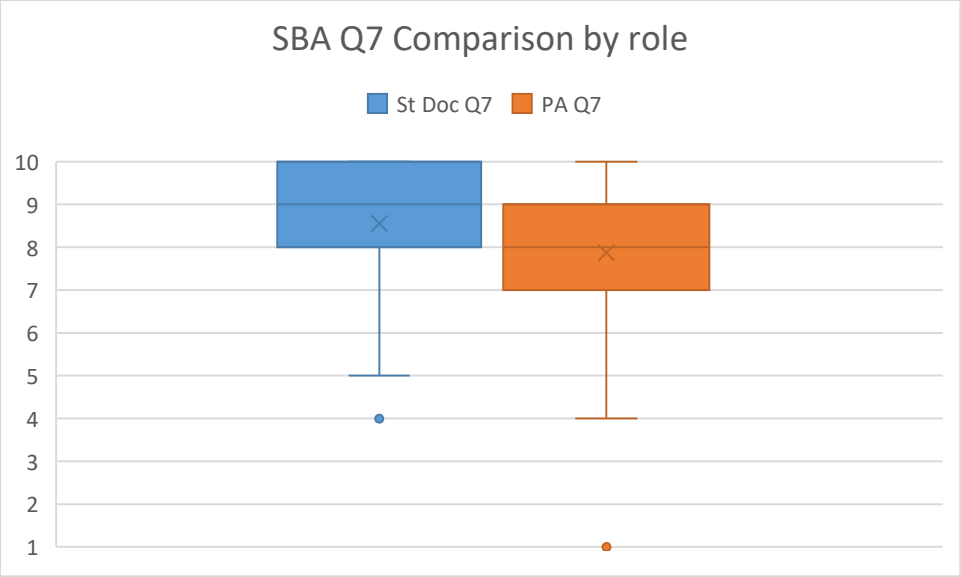
Table 29: Comparison of SBA data for difference between roles

Question	Levene's test significance	t value	Two-sided p result	Bootstrap
1	0.45	2.725	0.007	0.018
2	0.003	1.402	0.162	0.194
3	0.471	-0.162	0.871	0.867
4	0.114	1.414	0.158	0.167
5	0.735	-1.613	0.108	0.102
6	0.009	0.325	0.745	0.762
7	0.115	3.842	<0.001	0.002
8	0.450	2.697	0.007	0.008
9	0.260	1.568	0.118	0.136
10	0.565	-0.199	0.842	0.836

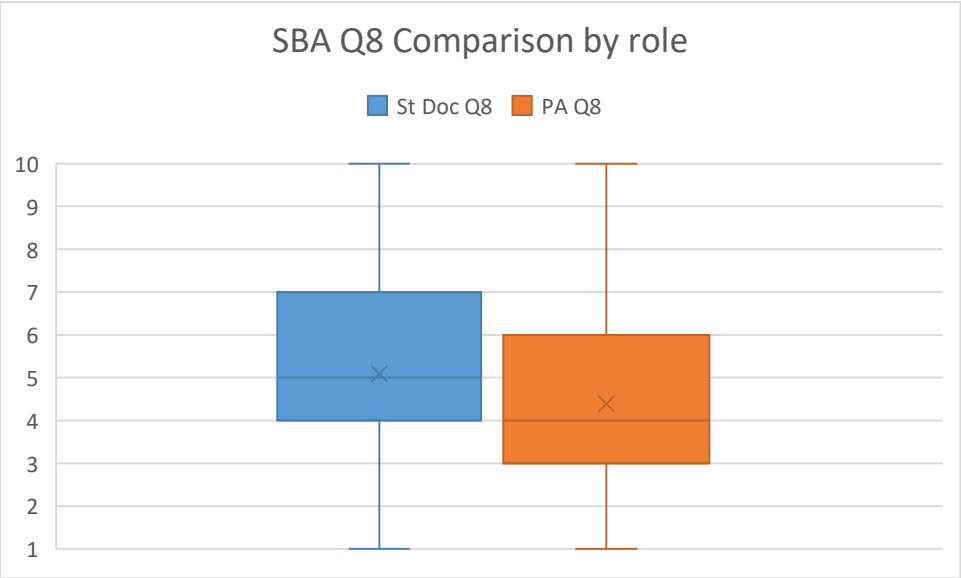
These results show a statistically significant difference in the answers to questions 1, 7 and 8. To explore this further these data sets were plotted as box and whisker charts together.



Whilst both groups felt their thinking was affected by being in an assessment the tendency was slightly lower for the student physician associate group compared to the medical student group.



Both groups show a strong tendency toward pattern recognition in their thinking processes for SBAs, however the medical student result is slightly stronger than the student physician associates.

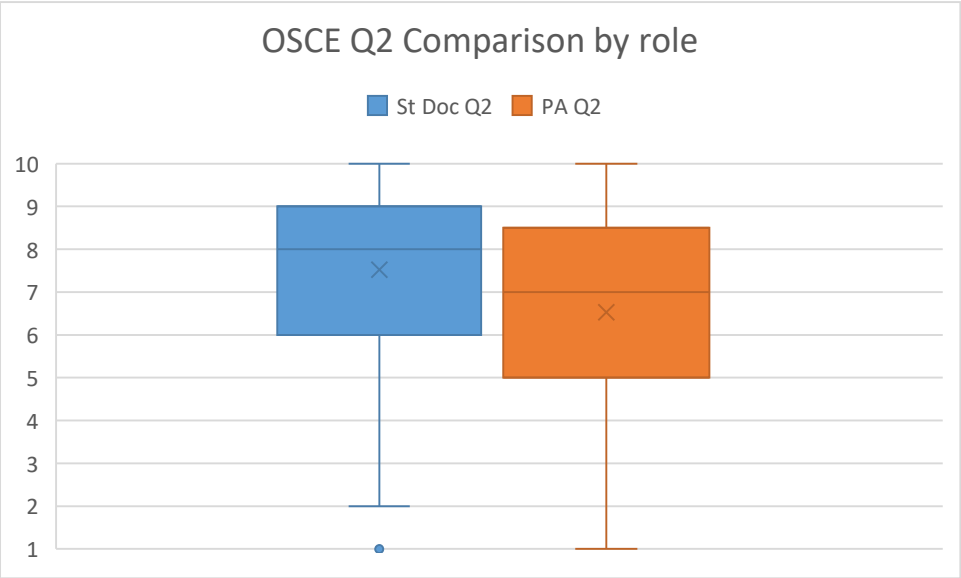


In both groups there was a tendency away from coming up with answers without understanding why, however this was slightly lower in the student physician associate groups i.e. they felt this occurred even less in their case.

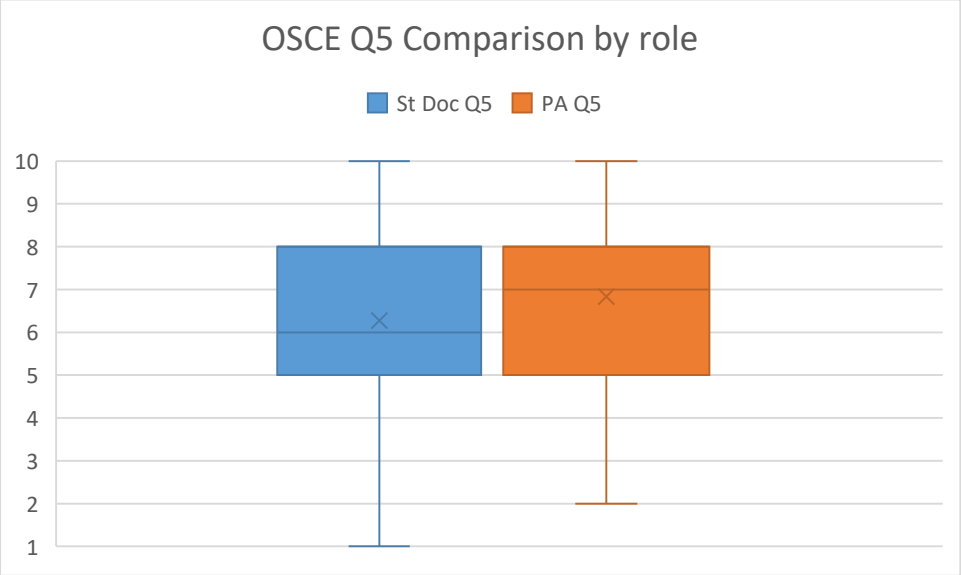
Table 30: Comparison of OSCE data for difference between roles

Question	Levene's test significance	t value	Two-sided p result	Bootstrap
1	0.366	1.892	0.60	0.041
2	0.039	3.310	0.001	0.003
3	0.869	-1.510	0.132	0.160
4	0.814	-1.265	0.207	0.228
5	0.236	-2.119	0.035	0.048
6	<0.001	0.972	0.332	0.403
7	0.005	2.828	0.005	0.018
8	0.156	-0.418	0.676	0.672
9	0.275	-1.340	0.181	0.180
10	0.407	-1.697	0.091	0.083

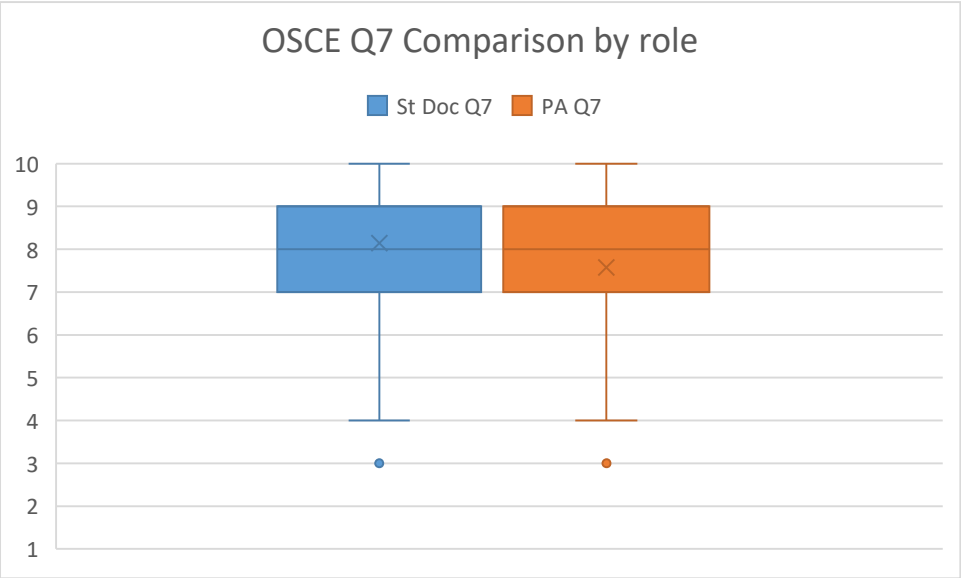
These results show a statistically significant difference in the answers to questions 2, 5 and 7. To explore this further these data sets were plotted as box and whisker charts together.



We can see from this that whilst both groups have a tendency toward seeing the demographics as important for decision-making, there is a slightly higher tendency amongst medical students.



In question 5 the spread of responses is almost identical; however the medical students have a lower mean and median than the student physician associates. This suggests that the medical students have a slightly lower tendency to formulate a list of answers before eliminating them one by one.



In this case the distribution and median are identical however the means differ. This suggests that there is a slightly higher tendency toward pattern recognition in medical students, mirroring the pattern for OSCE, however the difference in the overall data is minimal.

Overall, there is little difference between the two groups of clinical students in the results. Both have a high tendency toward pattern recognition in both OSCE and SBA, although this is slightly more pronounced amongst the medical students. This links with the slightly higher scoring for the importance of the demographics in their thinking seen in the medical student group as well. The other differences are relatively minor, though it is interesting that the student physician associates felt that the assessment itself affected their thinking less than the medical students. It does suggest that the longer course and, likely, greater number of assessments overall, for the medical student group may have a contribution in all these areas. It raises the question as to whether the tendency of medical students to use pattern recognition more strongly, associated with expert thinking, could be due to the increased level of exposure to assessments that they will have had in the longer course.

7.4.3 Impact of Gender in different examination types

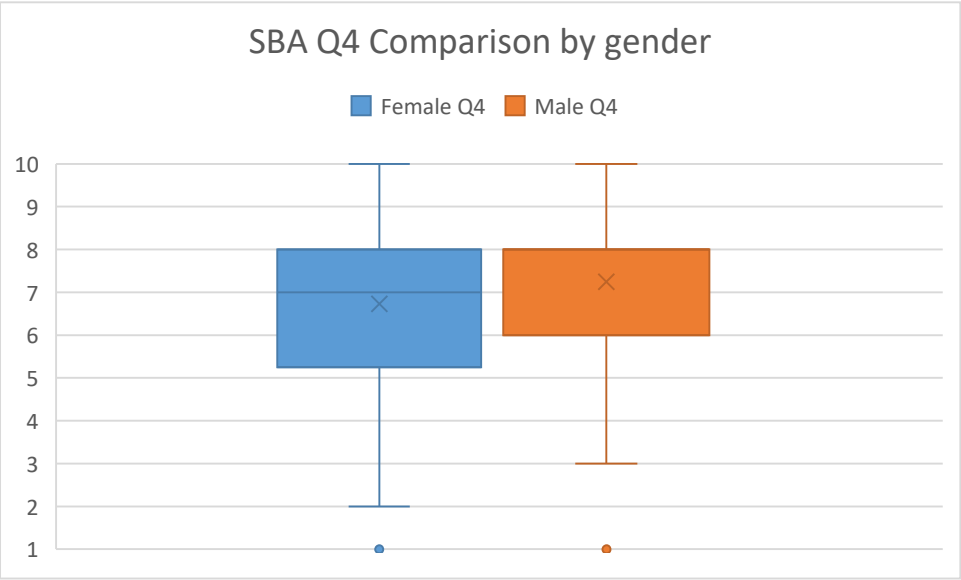
Analysis was undertaken to compare the separate datasets for the male and female groups. The test was looking for any statistically significant difference in the question responses between the two genders.

Any data point for “prefer not to say” or where the associated gender differed from that given at birth was removed from the data set. For OSCE analysis there were 94 male data points and 210 female data points. For SBA there were 74 male data points and 212 female data points.

Table 31: Comparison of SBA data for difference between gender groups

Question	Levene’s test significance	t value	Two-sided p result	Bootstrap
1	0.627	-1.617	0.107	0.157
2	0.350	-1.458	0.146	0.168
3	0.302	-1.521	0.129	0.151
4	0.547	1.975	0.049	0.049
5	0.557	0.444	0.658	0.661
6	0.988	0.410	0.682	0.653
7	0.521	1.378	0.169	0.159
8	0.546	0.251	0.802	0.826
9	0.646	1.759	0.080	0.066
10	0.868	0.887	0.376	0.391

These results show a statistically significant difference in the answers in only question 4 and this is right on the limit of statistical significance. To explore this further this data set was plotted as a box and whisker chart.

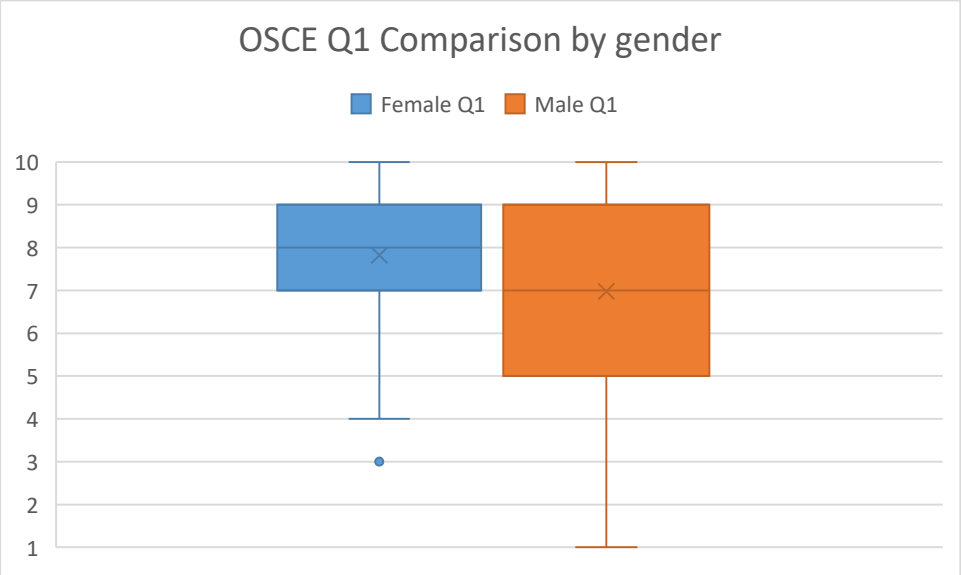


This data suggests that males have a slightly higher tendency toward early hypothesis formation which they then change as the question develops compared to females. The slightly narrower interquartile range and overall spread would support this conclusion.

Table 32: Comparison of OSCE data for difference between gender groups

Question	Levene's test significance	t value	Two-sided p result	Bootstrap
1	<0.001	-3.525	<0.001	0.007
2	0.786	-0.682	0.496	0.514
3	0.550	-0.748	0.455	0.448
4	0.789	0.051	0.959	0.961
5	0.532	-0.046	0.963	0.968
6	0.922	0.836	0.404	0.405
7	0.585	0.174	0.862	0.852
8	0.004	1.071	0.285	0.314
9	0.202	1.122	0.263	0.271
10	0.097	-1.660	0.098	0.119

The OSCE results show a statistically significant difference in the answers in only question 1. To explore this further this data set was plotted as a box and whisker chart.



These results suggest that females tend to feel that being in an assessment affects their thinking more than males with a narrower interquartile range suggesting greater homogeneity in this group than the male group.

7.4.4 Impact of Ethnicity in different examination types

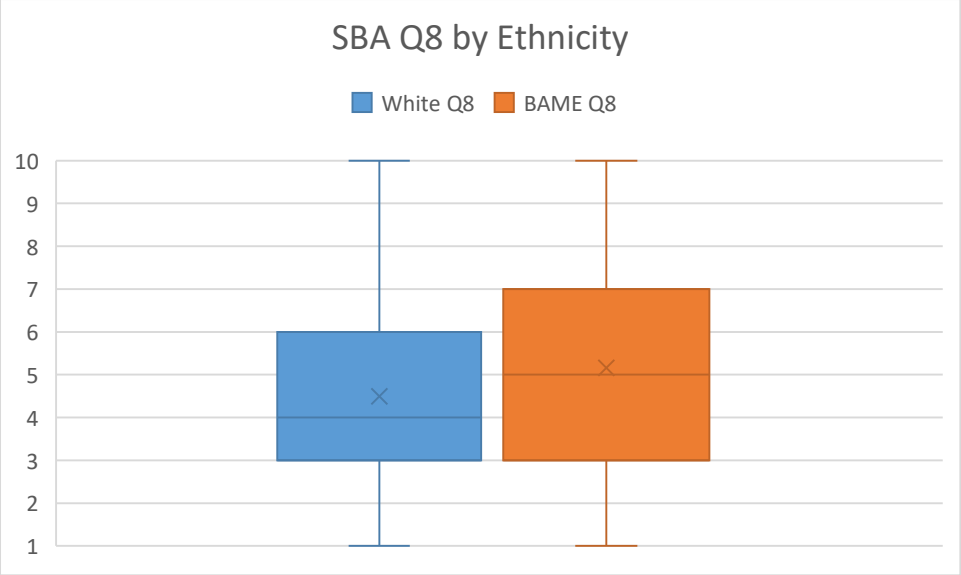
Analysis was undertaken to compare the separate datasets for the white and BAME groups as previously discussed. The test was looking for any statistically significant difference in the question responses between the two primary groups.

Any data point for “prefer not to say” was removed from the data set. For OSCE analysis there were 174 white data points and 127 BAME data points. For SBA there were 151 white data points and 134 BAME data points.

Table 33: Comparison of SBA data for difference between ethnicity groups

Question	Levene’s test significance	t value	Two-sided p result	Bootstrap
1	0.211	1.578	0.116	0.124
2	0.026	1.931	0.55	0.44
3	0.139	-0.617	0.538	0.559
4	0.990	-0.516	0.606	0.604
5	0.811	-1.516	0.131	0.153
6	0.631	-0.007	0.994	0.997
7	0.257	1.844	0.066	0.078
8	0.088	-2.607	0.010	0.013
9	0.354	-1.498	0.135	0.135
10	0.584	-0.859	0.391	0.423

These results show a statistically significant difference in the answers in only question 8. To explore this further this data set was plotted as a box and whisker chart.

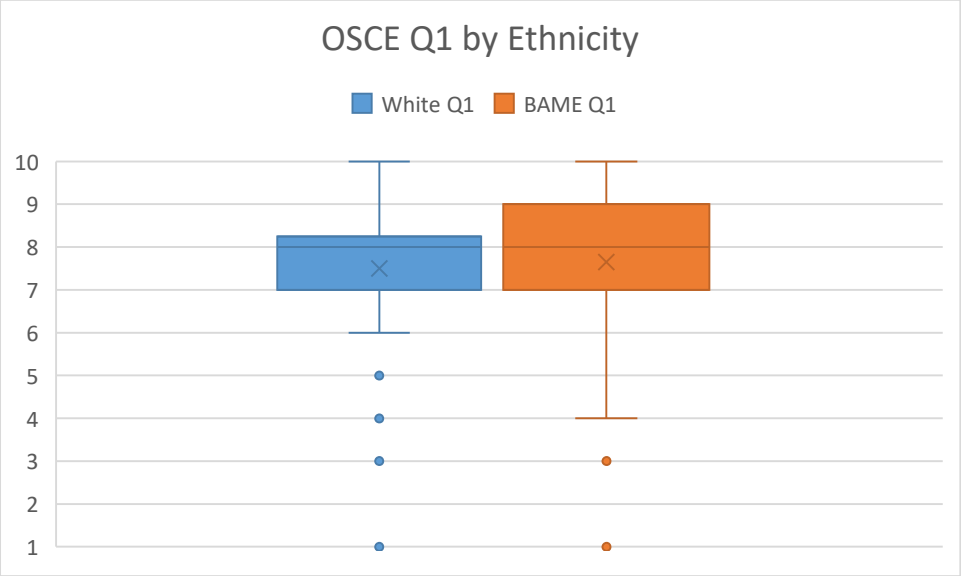


The comparison here shows that there is a slightly increased tendency toward coming up with answers but not quite knowing how you got there in the BAME group compared to the white group with a slightly wider interquartile range suggesting slightly less homogeneity in this group.

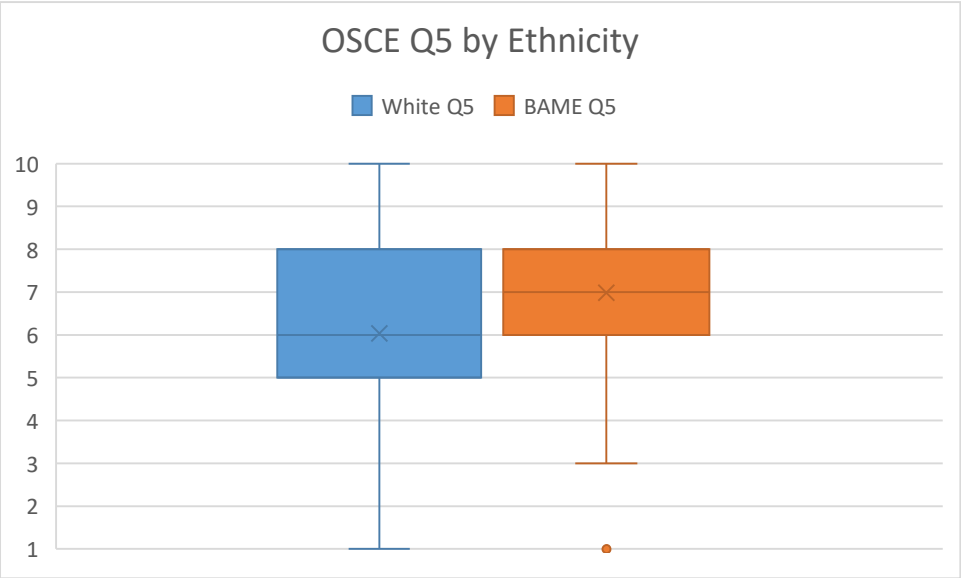
Table 34: Comparison of OSCE data for difference between ethnicity groups

Question	Levene's test significance	t value	Two-sided p result	Bootstrap
1	0.231	-2.053	0.041	0.049
2	0.070	-.0653	0.514	0.506
3	0.655	1.056	0.292	0.266
4	0.879	-1.884	0.061	0.64
5	0.980	-4.003	<0.001	0.002
6	0.048	1.280	0.202	0.222
7	0.010	-0.189	0.850	0.859
8	0.006	-4.396	<0.001	<0.001
9	0.029	-2.704	0.007	0.005
10	0.230	-0.854	0.394	0.410

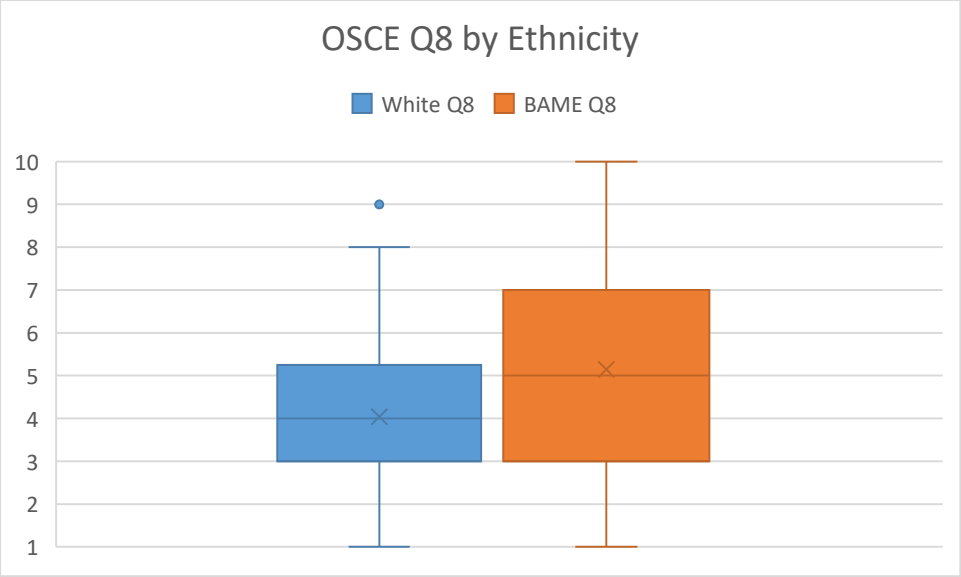
These results show a statistically significant difference in the answers to questions 1,5,8 and 9. To explore this further these data sets were plotted as box and whisker charts together.



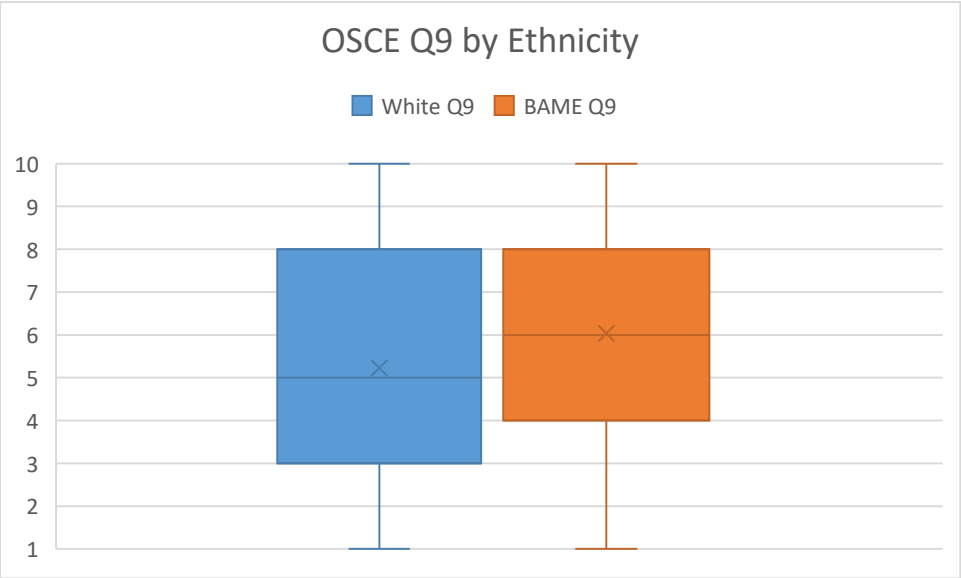
These results show a slightly higher tendency for BAME students to feel that the OSCE affects the way they think with a wider interquartile spread upwards on the scale and slightly higher mean. It is worth noting however that the overall data spread is greater for this group with the white students showing more homogeneity in their responses.



Question 5 answers suggest a slightly higher tendency of BAME candidates to formulate a list of answers then eliminate them one by one. It is notable that the inter-quartile range for the BAME group is narrower than the white group and all sits above the midpoint of the range. The white group have a wider overall spread and thus less homogeneity for this question than the BAME group.



Question 8 is about getting to answers and not being sure how you got there. The very wide interquartile range for the BAME students suggests less homogeneity of response, whereas the white candidates show a lower tendency toward this.



Given the wide interquartile ranges for both groups and the overall spread of results it is difficult to determine whether, other than the means, there is any truly significant difference between the two groups in this data, with both interquartile ranges around the midpoint of the data.

7.5 Quantitative Results Summary

7.5.1 The overall results.

The quantitative results can be broken down into the overall results and then the group comparators. The comparison groupings are designed to allow consideration of any statistically significant difference in the means between different groupings within the results, specifically the student clinical role, gender and ethnicity differences.

The overall results suggest a significant impact of contextual factors upon student decision-making, with both the context of being in an examination (student situational factors) and the contextual factors in the question itself (patient demographic and site – question situational factors) having clear importance and impact on decision-making. These responses also fit with the strong tendency toward pattern recognition. It is notable that this is the type of decision-making that is considered to be associated with expert thinking, therefore it raises the interesting question as to whether our assessment formats are driving this as a "gaming" approach, rather than a development of the individuals clinical reasoning. We must acknowledge the possible issue of this being at odds with students declaring that

they had a tendency toward trying to keep an open mind, which may link to some of the comparative findings.

There is a clear tendency toward early hypothesis formation and the results would suggest that students adopt either an elimination or repeated change approach to the decision process, fitting with abductive decision-making, but then individual tendency toward a more inductive or deductive end of the decision-making continuum. Students generally do not associate strongly with getting answers with no clear understanding of how, but they do associate with good-enough answers rather than perfect ones. The data suggests that this is clearly quite individualised as shown by the wide spread of answers and interquartile range.

Finally, students are aware of biases and how these may affect them however the level to which they understand biases is unclear, and whether they understand them in terms of their typical description, as heuristics can often be mislabelled biases.

7.5.2 The comparative analyses

The comparative analyses demonstrate some differences in certain groups and exam types. It is important that we are realistic in our interpretation of the results. Even where a statistically significant difference in the means exists it is only greater than one Likert data item point in a single question, and for a single comparator (white students vs BAME students in OSCE for Q8). Whilst for statistical analysis there is disagreement regarding whether or not the scale could be considered equal in gaps between points, the differences in the means are overall less than one single point across the scale, suggesting that even where differences occur these are small and therefore definitive conclusion cannot easily be drawn.

The areas of more interest are where the interquartile range is much wider for one group than another, such as with OSCE Q1 related to gender and OSCE Q8 related to ethnicity. These would suggest much less homogeneity of view in these groups around the answers, with the potential to introduce impact from a difference of thinking approach.

7.5.3 A note on the randomised groups

In 5.4.1 it was noted that a single question (9) had a difference in clinical setting to see if this influenced the answers, and in 6.4.1 it was noted how randomisation was undertaken. The results of this showed that only 26 of the 291 respondents chose to keep a patient at home

with 93% of group A (Emergency Department) and 90% of group B (General Practice) selecting an admission deposition for the patient. It was therefore decided not to pursue this further however this may provide useful data for future research where the same question might be used with experienced Emergency Department clinicians or General Practitioners.

7.6 Summary

On their own the quantitative analysis allows us to draw several clear conclusions. Pattern recognition linked to key contextual factors in the questions indicates clinical reasoning characterised by induction as the primary decision-making process utilised by clinical students in SBA and OSCE. Early hypothesis formation is common, although the data shows slight differences between question types, with SBA slightly stronger on pattern recognition and OSCE showing a slightly stronger tendency toward keeping an open mind.

Whilst the comparative analyses show some significant differences between the means of specific questions the differences overall are relatively small. If we consider this in relation to some of the issues raised around performance by different groups in assessments, e.g. the data from the GMC, it is unlikely based on these findings that this is due to different thinking approaches.

These results will be considered alongside the qualitative findings in chapter 8 to draw more meaningful conclusions.

Chapter 8 – Qualitative Results

8.1 Introduction and Chapter Summary

This chapter sets out the results from the qualitative data collection undertaken at the three study sites. The analyses are conducted in line with the methodology and method as set out in chapters 3 and 6.

The chapter will initially set out the core demographic makeup of the focus groups, the coding approach taken, and the approach to assure data saturation as it pertains to qualitative data. The chapter will then go on to consider the results, related directly to the different elements of the concept framework from chapter 2, and broken down into four main data sections – the cognitive continuum, external context, in test context and closure.

There will be a brief summary at the end of the chapter. Detailed discussion of the implications of these findings will not be included as this is a mixed methods study and conclusions should be drawn from the combined data as set out in chapter 9.

8.2 Demographics and Approach to Analysis

The sample for analysis consisted of seven focus groups collected at the three sites, with a mixture of medical students and student physician associates, as shown in table 35. All focus groups were conducted by me and lasted between 57 and 69 minutes. Each group has been assigned a number corresponding to the order in which they were conducted. Each focus group lasted between fifty and seventy minutes.

Table 35: Focus Group make up.

Site	Group Detail	Group Number	Medical Students	Physician Associate Students
University 2	Focus Group 1	1	5	13
University 1	Focus group 2	2	[redacted]	6
	Focus group 3	3	2	[redacted]
	Focus Group 4	4	3	
	Focus Group 5	5	4	
	Focus Group 6	6	4	
University 3	Focus Group 7	7	[redacted]	9
[redacted]			**18**	**28**

As part of the data collection at each focus group, baseline demographic data was captured to give a clear picture of the group involved.

- 64.4% of the students were student physician associates compared to 35.6% medical students.
- 84.4% were female with 15.6% male with all reporting that this was the gender that they associated with at birth.
- 53.3% classed themselves as white British whilst the remaining 46.7% were made up of a range of BAME ethnic groups.

It is worth noting that Physician Associate courses are heavily female dominated, and this may well have an impact on the gender difference in light of the distribution of student types.

All the groups were single clinical role specific (medical student or physician associate student) except for group one which was mixed. Quotes from focus group one participants are therefore labelled as "student" only as determination of the role of the responding individual could not be determined from the audio recording.

A thematic analysis process was undertaken as set out in chapter three (p52). This was a deductive thematic analysis process, as the primary questions used in both analyses set the primary areas for exploration. The thematic analysis process showed that the emerging subthemes sat underneath four overarching themes which will be used for the presentation of the results.

- Cognitive continuum – data pertaining to the reasoning process itself and its position along the cognitive continuum from inductive to deductive.
- External context – data pertaining to contextual factors that were not related to the examination questions themselves including elements such as the situation of being in an examination, context specific to examination types and student specific elements such as affect.
- In test context – data pertaining to context elements within the test questions themselves such as demographics, clinical setting and ethnicity.
- Closure – data pertaining to satisficing and finalising thinking.

Figure 10 demonstrates how these four areas relate to the originally developed concept framework and therefore relate to the cognitive process as a whole.

Despite "external context" being the first part of the concept framework it is the cognitive continuum that will be considered first in this discussion of the results. The reason for this is that the cognitive process provides the central point from which context and closure can both be discussed and their impact on the cognitive process considered. Metacognition, the way students think about thinking, will be considered within each of the sections as it applies to each of these four overarching themes.

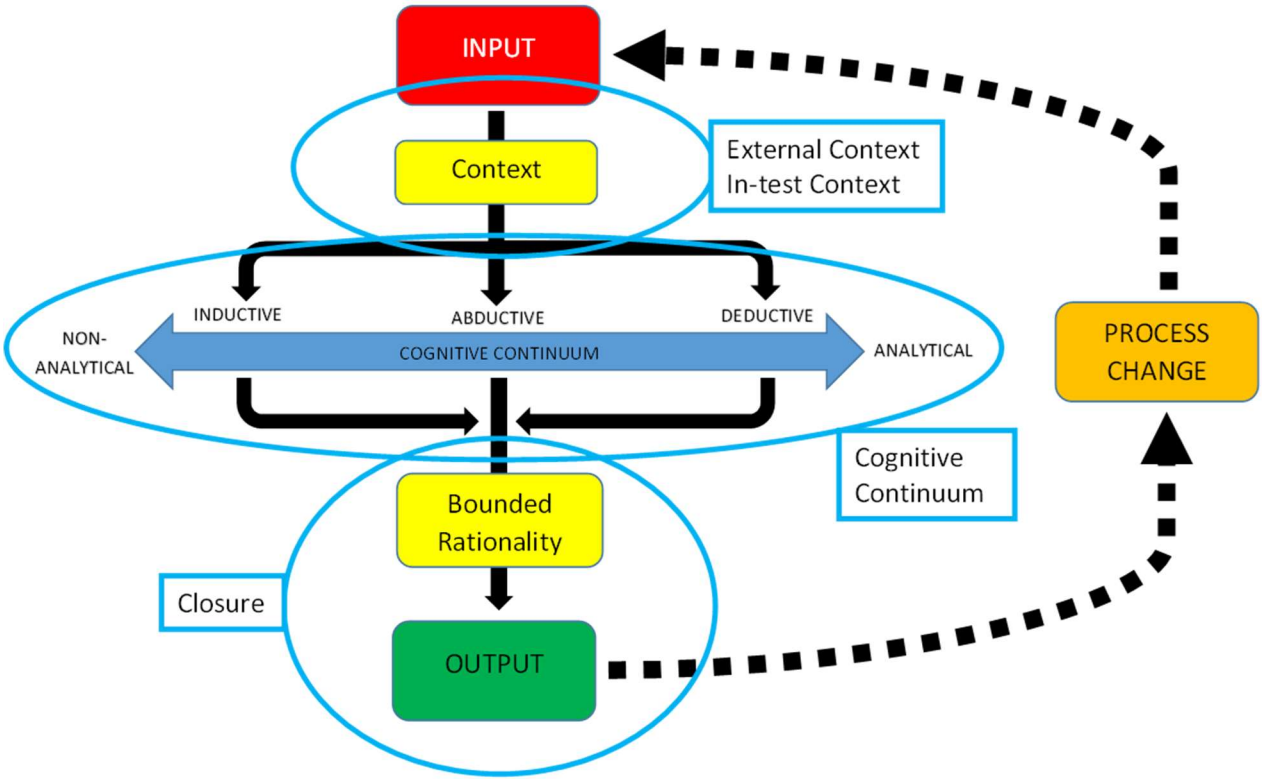


Figure 10: Concept framework demonstrating the four overarching themes.

The seven different focus groups were analysed using NVIVO release 1.7.1. and produced a total of 89 codes underneath which sat a total of 863 individual data points. In this qualitative analysis a data point constitutes a single section of the data pertaining to any specific code. The final code map and hierarchy is included as appendix 11 and this is colour coded to facilitate clarity regarding the stage of the process at which each code was generated.

As discussed in chapter 3 (p52), the number of new codes generated at each stage of the process was recorded for the purposes of validity and are graphically represented in figure 11. The initial codes generated from the pilot analysis are included as these formed the initial framework from which the rest of the analysis progressed. The majority of codes were generated from the first and second focus groups in the primary study data, with very small numbers generated thereafter, suggesting that the data produced is likely to be representative and valid.

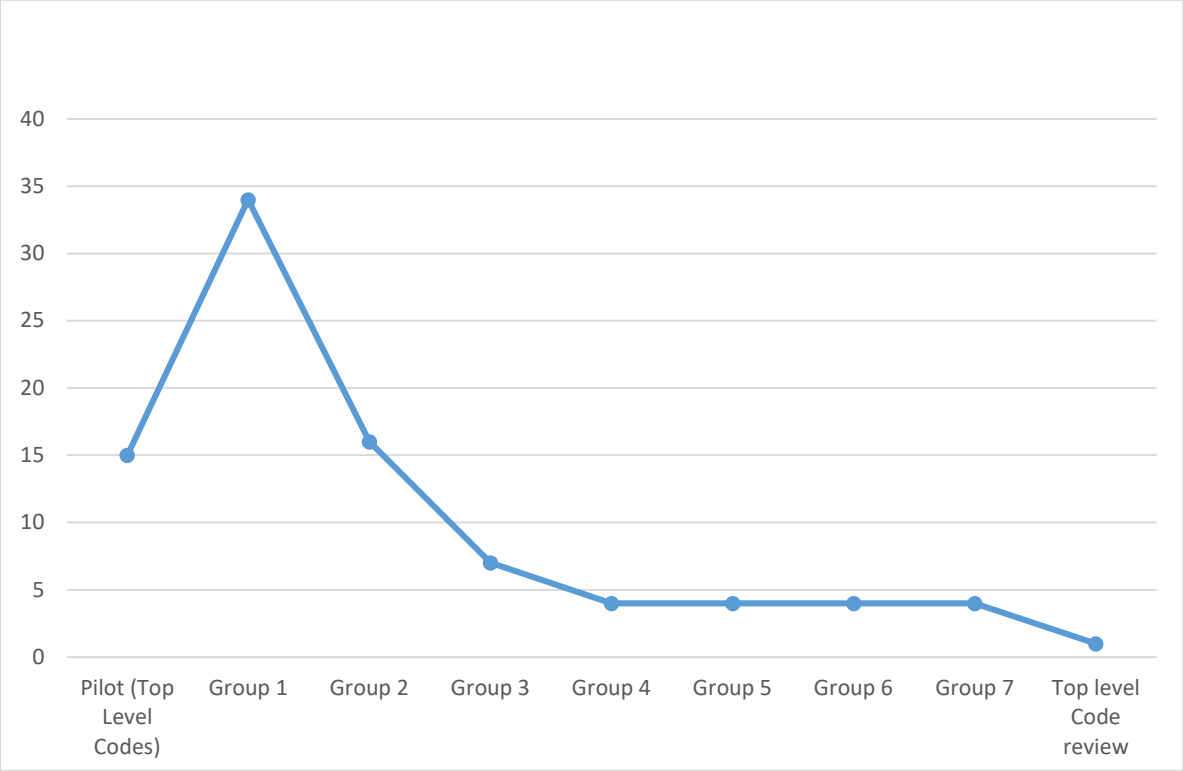


Figure 11: Number of new codes generated on review of each of the focus groups

Whilst the pilot focus group was not included in the final data set, nor revisited or reviewed against the later generated codes it is included in the graph (dimensions) for clarity and transparency.

8.3 Cognitive Continuum Results

Cognitive continuum relates to data that is concerned with the clinical reasoning process, within the assessments, along the continuum of inductive (pattern recognition) to deductive, as set out in the concept framework produced in chapter 2. The section of the concept framework relevant to this discussion is shown in figure 12. As with all of the results

the main subthemes will be discussed individually and then metacognitive elements relating to it will be considered.

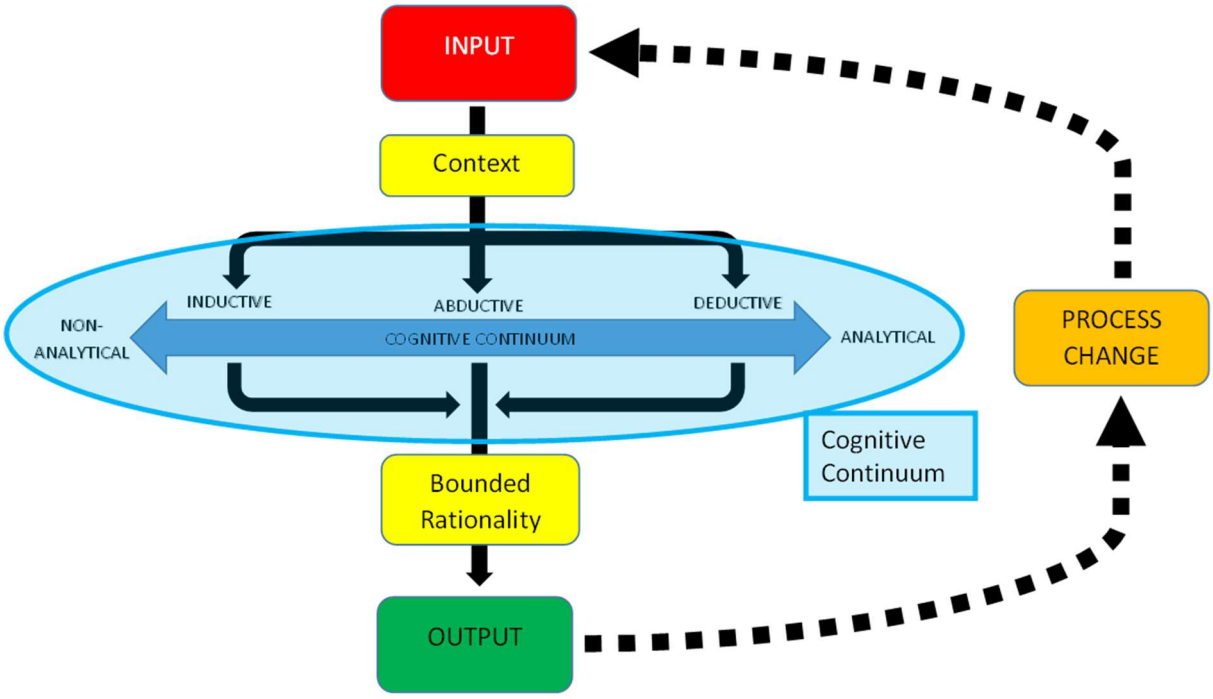


Figure 12: the cognitive continuum element of the cognitive model

8.3.1 Early Hypothesis Generation

Early hypothesis generation was articulated by students in relation to both exam types studied. In SBA this was particularly linked to pattern recognition from past practice questions as students immediately generated answers based upon their previous experience. Even where they did not immediately have a rapid, inductive, answer they would tend toward thinking of potential answers immediately they started reading.

> "You read sort of the first line in your head you're already thinking of what this stem normally leads to, with like, with an answer. So, I think as soon as you read the question, I'm sort of starting to form differentials and thinking what, like, it could be this, it could be that" *Medical Student, Focus group 5*

SBA questions were also noted as being different, in that the answers themselves present the differential diagnosis, and thus any additional approach to hypothesis formation was of little value. Students recognised that the answer had to be one of the five presented options, so some took the approach of immediately reading the answers and then going

back to review the stem to see which one fitted best. In essence this approach uses the answers as a surrogate for early hypothesis formation. It was also noted that there was sometimes the potential with this approach that it might introduce answers to the individual that otherwise may not have been considered as well as enabling a "rule out" approach, a more deductive approach, than would otherwise have occurred. This was also noted when the question was particularly difficult, with students feeling they needed to use the answers to understand the stem and then answer as best they could. Some students noted that whilst they would endeavour not to look at the answers first, they often found this impossible or that previous experience of hypothesising, and then finding their own answer was not present, was both demoralising and a waste of time. This approach was recognised to have some risk, with some students being aware that there was the potential to persuade themselves an answer was correct that they might otherwise not have considered. This is the bias of premature closure and students were aware of its potential to derail their attempts to enforce effective thinking strategies.

> "I always glance at the answers first just to see, right? This is a psych question or oh, this is a haematology question. So, I'm going in with that already mindset because it's got to be some of those. Whereas if you just start and then find the answers, you could have gone oh well, i thought it was anaemia and it's asking me like, none of these are the options. So, I always glance quickly at the answers and then read the question properly" *Student, Focus group 1*

In OSCE, the data suggested that the approach was slightly different with students using early hypothesis generation to focus their thinking on the questions that they would want to ask so as to maximise their chances of doing well. The differentials generated could then be ruled in and out through questioning and based on the information presented.

> "you're coming up with you, some differentials as I've said, otherwise you wouldn't be able to target any of your questions, you wouldn't be able to rule anything in or out if you didn't have some hypotheses to begin with" *Physician Associate Student, Focus group 2*

There were some subtleties to the cognitive approach that came through in the data. The approach that favoured the early hypothesis could lead to clear cognitive unease when the student found themselves unable to produce a hypothesis. This was noted particularly in the OSCE examination where the inability to go into the station with some initial thoughts was described as a "bad outcome" by one student. Another stated that in that situation you just

"picked one" (a hypothesis) and hoped that it was acceptable but, as described further below, trusted that the mark scheme would not overly penalise them for this approach.

> "if you get to a point when none of them feel like (the right one), there's, yeah, you just pick a random one because you're like I don't know which to choose. So, and then I think we were told that is probably like one mark so your just over it. So as long as you mentioned everything else then we're fine. But then again you never know." *Physician Associate Student, Focus group 7*

Students also noted that there was a difference in the approach, depending upon the decision that was needing to be made, specifically diagnostic decisions rather than management decisions. Diagnostic decisions were more associated with inductive reasoning, for example described by some as a "gut feeling", whilst others stated that this was more of an elimination and inclusion process. With management decisions some students described a rapid decision process whereas others saw this more as a deductive step wise process. This shows that the cyclical process that occurs as the decision changes (I have a diagnosis and now I must return to the beginning to consider a management plan) can lead to different cognitive approaches. This links to the concept of abductive reasoning, however the difference in the data provided can lead different students more in one direction on the continuum or the other depending upon their own personal thinking style.

8.3.2 Preferred reasoning style

It is important to recognise that students expressed different preferred learning styles within the data, which can impact upon the reasoning approach.

> "I'm bad at rote learning. Flashcards, awful, can't do them…..a lot of my background is built on why do things happen and like why?.....So I think it's unusual for me to get to an answer and not know the pathway behind it just because otherwise I wouldn't remember that was the answer." *Medical Student, Focus group 6*

> "It's just going back to like what type of learner you are though? Because my entire exam was based on ANKI, so it's like flash cards. So, I'm very like word association. I see one word, my mind goes to another and like gives me the answer. Uh so I think my exam technique is much less understanding based." *Medical Student, Focus group 6*

It was clear that there was variation in the thinking in examination types. SBA examinations had strong preference for pattern recognition (inductive reasoning), however this came through less strongly in the data relating to OSCE. Nevertheless inductive reasoning was still present at a more dominant level than deductive reasoning in OSCE. Specific phrases stood out in the data regarding SBA thinking such as trigger, or buzz, words and that these then

would automatically trigger a clear association in the students' mind leading them to a particular outcome. Students recognised the potential advantage of this within the examination setting.

> "I find basically I'm waiting for them to say something that triggers "Ah that's what it is"" *Medical Student, Focus group 4*

> "Yeah, prostatitis where they say it's like a boggy prostate or craggy prostate for cancer, like it's very set words." *Physician Associate Student, Focus group 2*

The responses throughout the focus groups suggest that there is a preference for pattern recognition (inductive reasoning) in SBA questions, whilst for OSCE it is more mixed. However there are subtleties to this that need to be considered within the context of the examination types themselves which will be discussed below. It is notable that several students commented on the fact that pattern recognition was a natural part of reality and important in order to do their job properly.

> "I think medicine is pattern recognition really isn't it and that's why the more experienced you get the more a consultant will say oh yeah I've seen this however many times before so I know that this is, we should do this, or whatever and because I think it's the more you see it, the more things come up or the numbers come up or a patient presents in a certain way." *Medical Student, Focus group 4*

This observation fits directly with evidence that inductive reasoning is linked to expert thinking. It also supports the previous research evidence that medical students also undertake inductive reasoning, as opposed to simply a deductive reasoning process, just not perhaps as expertly.

8.3.3 Reasoning within SBA

The reasoning within the SBA examinations was found to be heavily inductive with the recognition of patterns in the question being the primary approach students took to reasoning their answers. This was linked to the use of practice question banks and a recognition that there are only so many ways that questions can be written about specific topics. Students recognised that best practice in writing SBA questions meant that they should only contain information that was directly relevant to answering them, and that they had to fit typical presentations, as to introduce excessively grey areas was likely to lead to questions that were not fit for purpose. This led to students recognising that in SBA questions a piece of information being present in the question must be relevant in the

decision at hand and thus will have to have an impact on the possible answer, a clearly inductive process.

> "there's only so much of a stem that you can put in, you can't put in a page or whatever so to get across a certain condition the person who's writing the exams, has to put some kind of pattern in or something. So, I think that's really important to just recognise them when they are there." *Student, Focus group 1*

The data emphasises the idea of trigger words and phrases as the mainstay of the student learning approach for these examinations and consequently leads to a self-fulfilling prophecy for the use of inductive reasoning when answering this question type. Students commented on the fact that the teaching that they receive whilst in education, along with the process of revision, was designed to lead to pattern recognition. The fact that much of the teaching they have is condition specific, with the classic presentations of those conditions, will push them toward this process. The use of revision resources, specifically question banks, was commonly associated with the process of inductive reasoning via patterns with some students citing the fact that having completed large numbers of questions through these sources the process of pattern recognition was inevitable, with the point raised previously about there only being so many ways a question can be asked coming through strongly in the data. Some felt that the selection process for medical school was also focussed on pattern recognition and therefore it was inevitable that this would end up as a dominant thinking process in successful applicants. The UCAT (United Kingdom Clinical Aptitude Test) test was highlighted as having components heavy in pattern recognition as were the core examinations at ages sixteen and eighteen which schools utilise as their primary entry requirements. Whilst the pedagogical approach was primarily cited as driving this thinking behaviour it is notable that some students reflected it as being part of the human condition and therefore inevitable that it would form part of the cognitive process.

> "I just think we're built for pattern recognition. Like you walk out, you see clouds in the sky you go ah, the sky's gonna rain. You've recognised the pattern. You think I'll get a coat otherwise I'll get wet. And that's like all of human survival." *Medical Student, Focus group 6*

8.3.4 Reasoning within OSCE

The OSCE examination was described as a performance by many of the students and there was a recognition that it differed significantly from the SBA papers, due to the need to draw

information out of the actor/patient, rather than being presented with the information directly. Despite this awareness pattern recognition was still strongly represented in the OSCE, but in different ways to that in the SBA.

There were some students who noted that they approached OSCEs in a very specific way, based on past experience and practice, so they would settle into following that previous pattern when the station seemed familiar. They commented that if those patterns were then thrown off by something that was said by the actor it could unsettle their thinking as they had become fixed to the previous pattern they had experienced. Similarly, others commented that there were only a limited number of different types of OSCE stations that were likely to come up, relating this back to the idea that the exam is designed to ensure that those passing would be safe clinicians. This also indicates that pattern recognition around OSCE station construct was present in the student thinking.

There were also some very task specific approaches to the OSCE related to the nature of the OSCE mark scheme itself. Students recognised the fact that to succeed the key was to accumulate points rather than there being a specific requirement to achieve a final diagnostic or therapeutic decision. This led to some behaviours that focussed on patterns related to points acquisition in order to maximise their chances of success.

> "It's more of a checklist. You've done your reflection, you get a point, you've done this, you get a point" *Physician Associate Student, Focus group 2*

Some saw deductive reasoning as advantageous in the OSCE scenario and actually sought to control their reasoning approach in order to prevent the risk of cognitive biases occurring, particularly anchoring and confirmation bias.

> "For OSCE it's definitely about keeping an open mind until I've got the whole picture and, ok it seems like that fits that best and maybe have something else in the back of my mind that it could be." *Medical Student, Focus group 5*

Despite some evidence for this deductive approach the majority of the data still demonstrated that students focussed on trigger words or information constructs as important in OSCE, with specific scenario descriptors leading to them taking a pattern driven approach. The nature of the OSCE was such that students often talked about asking vague questions, or patterns of questions, until the relevant phrase came out and led to their "eureka" moment and they could then follow a pattern driven internal script to try and

deliver optimal performance. Students recognised that this approach was a higher risk for bias, specifically anchoring and confirmation bias, evidencing a recognition of these phenomena in OSCE, which is natural due to the nature of the exam needing a fact-finding approach, rather than an SBA where all of the information is already presented to the student to consider.

8.3.5 Cognitive Continuum Conclusion

The overall picture regarding the reasoning style on the cognitive continuum is that, in both SBA and OSCE examinations, students utilise thinking that is focussed on the inductive end of the continuum in the majority of cases. The evidence would suggest that overall it would be more appropriate to describe the thinking process as abductive, particularly in OSCE where information is gathered over a period of time. The concept of the cognitive continuum is seen clearly in the focus group data with general preferences tempered by evidence of different approaches, depending upon the information presented either in the SBA stem, or the OSCE introduction.

8.4 External Context Results

External context relates to data that is specific to the situation of the student being within an assessment setting and how these have general, and exam type specific, effects on the thinking process. This forms a key part of the initial process. Its position within the concept framework is illustrated in figure 13.

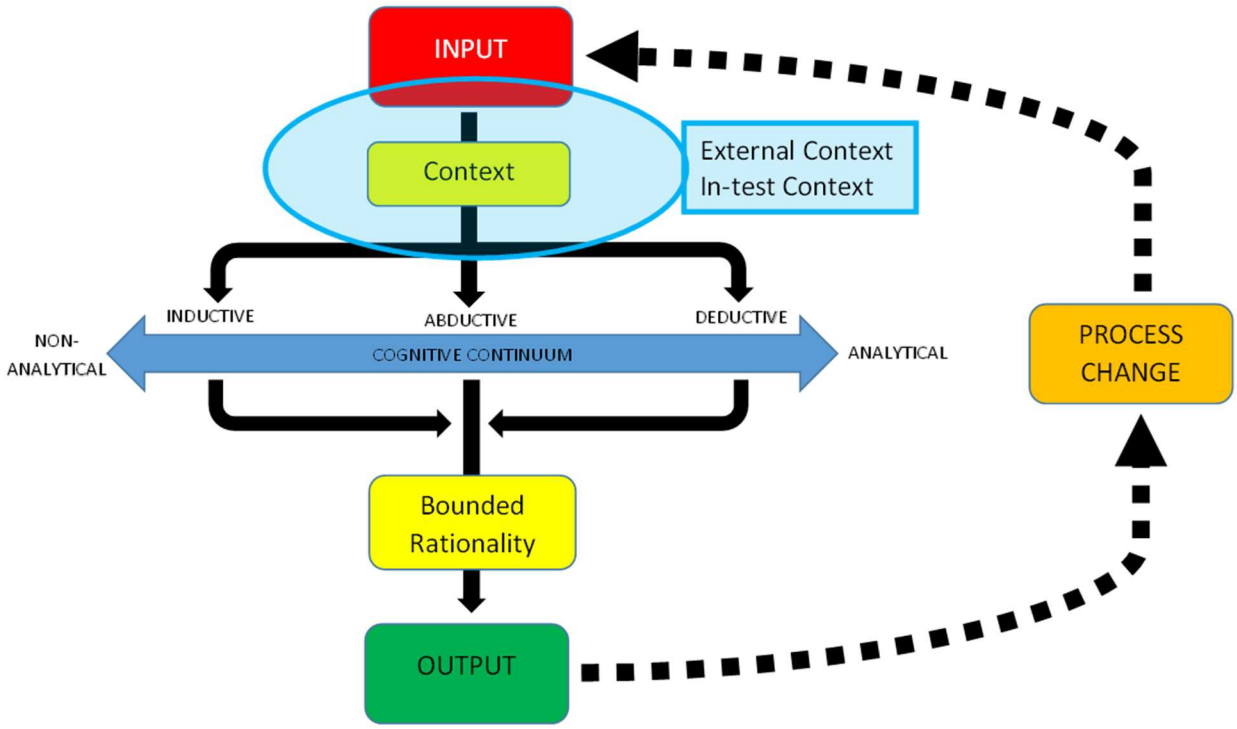


Figure 13: The position of context in the cognitive model

8.4.1 Student affect

Students were clear that the way that they felt affected their thinking in examinations. This went back to hours before the examination, through aspects such as the challenge of effective pre-assessment sleep, and the concerns that this caused regarding the potential for concentration lapses, as well as general anxiety. There were a number of comments that were more specific to these being final year students, around the knowledge that their career depended upon the outcome of these examinations, and this led to a feeling of added pressure. In terms of the anxiety affecting thinking, the idea that nerves would lead to students struggling to think clearly, or perhaps more accurately struggling to close their decision, came through in the data.

> "I think if you're nervous at the start, I think there's a higher probability, you'll just second guess yourself. I think if you go in with the mindset that you know your stuff and it's just a case of you just need to answer the stuff, answer the questions confidently, you're a bit less to sort of second guess yourself." *Medical Student, Focus group 3*

There was a longitudinal element of change to this thinking in the data with students talking about how exam stress, time pressure and fatigue led to changes in thinking; this will also be covered under the outcomes theme, with the SBA style assessment being most notable for this. Stress in general was an issue for students. The concept of eustress leading to

enhanced performance is well understood, however distress can lead to worsening cognition and outcomes in assessments. Students struggled to articulate exactly how the stress affected them but recognised the impact it could have with terms like “nothing was going on in my brain”, “impact my memory” and similar phrases indicating a general impact with adverse effects on the reasoning process.

> “I think generally anything that can make you feel more anxious, will affect the way you answer questions. So really, I wouldn't be able to say in what kind of way, but kind of any kind of external thing that's going to add to anxiety. I think that's for me. That's the biggest feeling that comes in to doing exams. I wouldn't say necessarily that like feeling sad or anything like that would come into it, it's more, just the level of anxiety on the day and other things that have played into that.” *Physician Associate Student, Focus group 2*

An interesting observation was the effect of question order and its link to stress in the examination which was commented upon as an issue in both SBA and OSCE examinations. Students commented that if the questions, or stations, started off with higher difficulty then this led to higher anxiety which then impacted their performance. There was data suggesting that students could start to feel that the exam difficulty was leading to an inevitable outcome of poor performance and a mental inertia, whilst others commented that once they had got into a rhythm with questions they could then go back and find these difficult questions easier to tackle, due to a degree of cognitive ease now having developed. This led to some utilising deliberate metacognitive approaches to managing this issue. The nature of the OSCE examination, and the inability of students to return to previous stations and review performance, also led to a similar issue. Students commented how certain stations were more difficult to “let go of” as they progressed to the next one with a risk of their performance being affected as they could not fully focus on the new scenario that was now in front of them.

> “If you start off on a really hard OSCE, I can imagine you'd be just very anxious. And yeah, not really in that relaxed state of mind that you need to be in for an OSCE because you, a lot of the time you, you just then would forget simple things because your just so bewildered from the last thing that happened that your just in that state of mind” *Medical Student, Focus group 4*

Some students commented that as the time pressure came more to bear upon them they become more confident in their answers, though perhaps that is not the appropriate way it might be interpreted. Clarification suggested that the word confidence was a surrogate for

students tending toward more inductive reasoning as the exam progressed and the time pressure increased. This demonstrates the clear impact of external context, in this case time pressure, on the student reasoning process.

8.4.2 Exam Setting

In both SBA and OSCE the setting of the examination was considered to have an effect on the students thinking.

SBA questions lend themselves to being taken online and therefore can be completed in any appropriate setting. Students noted that when at home it led to a sense of ease both due to the familiarity of the environment but also the fact that they had grown used to it from doing practice questions there. Being in an invigilated setting was therefore out of the comfort zone of many and thus could affect how they felt and their decision-making process. It was not however a uniform view as others stated that the exam setting enabled them to focus on the fact that this was an examination and therefore made them concentrate more compared to sitting at home where there were always distractions.

The other element that was noted was the other people present, through their own processes for managing their thinking, might leave prior to the end of the examination which could disrupt the thinking of the individual and lead to higher anxiety as they found that in seeing other individuals leave they became worried that this meant that they were performing poorly in a potentially easy examination.

> "Well like you were saying like, if you're like, if people are like, leaving the exam and you're still sat there, I mean that can make you increase your anxiety like why did they find it so easy?" *Physician Associate Student, Focus group 2*

For OSCE there were clear indications that the setting affected thinking initially. Some commented that being in individual rooms was better as you did not get the external noise that is usual when these examinations are undertaken in more open environments, with students separated by curtains or other dividers. Some noted that rooms similar to General Practice surgery rooms aided their cognitive ease, and thus had the potential to help their performance, as they were closer to the real clinical environment. Those students who commented regarding the setting for OSCE using the more common standard dividers noted that as the exam progressed the setting simply became that of an OSCE and any relation to a specific clinical setting in which the scenario was based became somewhat irrelevant

(General Practice, Emergency Department etc). The issue of clinical setting is discussed more below.

8.4.3 Other examination and student factors

It is important to reflect that for this study formative, not summative, examinations were used for the primary data collection (apart from university 3 OSCE). Some data showed that students place much greater value on summative assessments rather than formative assessments and will alter their process accordingly.

> "Yeah, absolutely if it was end of year exam I would have gone back through maybe three times. With it being a formative I just thought…." *Physician Associate Student, Focus group 2*

SPLDs (Specific Learning Disabilities) was another area which came up as linked to the student affect within the examination. Students commented on how having the extra time led to an element of cognitive ease, noting that this was particularly the case in SBA examinations as there were seldom any adjustments for OSCE examinations. The view of students differed regarding OSCE with some regarding the reading time to be plentiful, as often the information provided outside of the station was quite limited, and others feeling that they were rushed. Similarly, there was evidence that knowing that they had an SPLD, and that they had to focus their concentration, led to some students feeling that they risked "missing the point" of a station with potential for performance detriment.

> "I have extra time in exams, which, sort of, well it's obviously helpful because then I have less time pressure than otherwise but in the OSCEs, when, you know, the scenario that you have to read isn't very long. You've got a full minute to read it. Even with me needing extra time, that is more than enough to read the scenario and to understand what you should be doing, what you're about to approach. But even then, with more than enough time I think I was so fixated on the fact that I had to read it and understand what I was doing that I didn't even take in what I was meant to be doing and twice I was told would you like to read the question again" *Medical Student, Focus group 3*

Although students with dyslexia felt that the additional time provided ensured that they had adequate time in the examinations, they still noted the challenge that some of the complex medical terminology provided relative to understanding the word in plain English. This also linked this to their own learning, with one student noting that they would tend to learn by writing in plain English rather than using medical terms, e.g. swallowing difficulty rather

than dysphagia. The data suggests that SPLDs can impact on the reasoning process within the examinations even where appropriate mitigating approaches are put in place.

8.4.4 Task specific thinking

The various different external contextual factors led to students adopting task specific metacognitive management approaches to controlling their thinking.

In SBA examinations students recognised the structure of the question and the implications that these must have. The awareness that all of the information contained within the question must be relevant led to focussed thinking, which inevitably leads toward inductive processes as previously discussed. Some students described specific strategies in the examination to enable them to target questions for review after an initial attempt to try and reduce time pressure such as highlighting questions in specific colours to target the ones they wanted to go back to and review, or others answering all the ones that they could do confidently before going back and attempting the ones that they were more uncertain about.

> "I'll try and kind of pick because I know that I'm more high adrenaline at the start of an exam, I'll go through and do the questions I can do quickly first because I know that they're the ones where I'm at least likely to make the silly mistake and I can kind of use the pressure up and then once I've kind of settled in and I feel calmer I'll then go through and do the like more complicated ones." *Physician Associate Student, Focus group 7*

Others discussed the equivalent of "reason aloud" strategies to try and manage their thinking and control initial nerves whilst they felt their way into the examination itself. Some talked about the risk of talking themselves out of answers, whilst others were clear that this was why they did only limited review and then left the examination, trusting their instincts regarding exam performance.

For OSCE, much of the task specific metacognition was focussed on the examination itself and the way that the mark scheme was constructed. Additional approaches surrounded the script that was available to the OSCE patient. The recognition that, whether patient or actor, they were working from a set script led to students looking for cues linked to this. The OSCE actor having to search the script for information, clearly being unaware of an answer, or otherwise indicating lack of knowledge in the area under exploration (such as through their body language) would lead students to change their questioning plan to allow them to

narrow down their thinking. Some indicated looking for clues as subtle as changes in the way an individual patient answered a question to indicate a need to rethink their strategic questioning approach.

> "When the patient is like, saying something that just, just seems completely unnatural you can kind of see them switch and they go, "oh yes I've been having" and you go yeah, a bit weird. And then you can like latch onto it or something that just seems quite out of character from the way they've been speaking. You can kind of notice what is scripted I suppose and when they look at their paper and you sort of think Ah." *Medical Student, Focus group 5*

8.4.5 Assessment versus the real world

Students were very clear in the fact that assessments were not the real world and that this influenced the way they approached their thinking, as well as having concerns about the potential risks that this could present when they were qualified.

As discussed previously students recognised that SBA questions would contain the information that was required and not elaborate, but this also means that all of the information is relevant. They noted that in a real clinical encounter the amount of information provided by the patient might be extremely limited but that they would be able to extract that information through careful questioning.

> "I think the problem with SBAs is in life you can continually ask. So if you're, if you're concerned about something and somebody's presented or something, you can go down that line of questioning whereas in a written exam you're given, like you're given a very short kind of synopsis of what they've come in with and what the problem is. And on that you have to make like a very snap judgment which you don't do that in real life" *Medical Student, Focus group 6*

The recognition that patients do not present with clear trigger words or prompts, in most cases, also featured in students concerns about the relevance of the thinking processes developed for assessments. The fact that pattern recognition dominates thinking for SBA questions was seen as potentially problematic in the context of seeing patients for real due to the risk it might pose to the introduction of biases in clinical practice, especially premature closure, confirmation and anchoring biases.

The time limitation that is applied to OSCE examinations was also felt to be unrepresentative of the reality of working clinically i.e. the ability to ask more questions at an appropriate point rather than having the time run out prior to having that opportunity.

"….in practice, you'd never walk out of a room and go shit, I bet that's ectopic pregnancy. oh well, I'll move on to the next patient. You'll go back and say I'm really sorry I've just, you know, just remembered something that I think we ought to consider, it's really important to rule in or out, do you mind if I ask you some more questions?" *Medical Student, Focus group 6*

8.4.6 External context conclusion

The data supported that students recognised that they had to have a specific way of decision-making for examinations which was driven by the needs of the assessment and influenced by the nature of the process that they had to undertake. Their own personal affect in the examination, along with their recognition of their own decision-making processes, led to specific ways of approaching SBA and OSCE examinations. Students recognised that these metacognitive approaches, or more accurately the metacognitive approaches they were forced into, would not then necessarily be directly applicable to working in a clinical environment on qualification.

8.5 In Test Context Results

As stated in the chapter introduction in test context refers to contextual elements contained within the test itself. Demographics and clinical setting were looked at in detail within this study. Figure 13 remains applicable for this section as in test context forms the initial part of the student thinking process and

8.5.1 Demographics

Demographics refers to specific features of an individual or population. In this study it focusses on the recorded demographic elements of patients within the assessments such as age, ethnicity, sex etc. These typically frame a question, whether SBA or OSCE, and provide a picture for the student to work with in their decision-making process. The primary focus of discussion around demographics was age and sex.

Demographics were strongly associated with the use of inductive reasoning by students in SBA examinations. This related to the concept of the "classic" patient and that certain specific demographics would automatically trigger recognition linked to the heuristics that students pick up through their training. Examples of this would be the typical patient for a myocardial infarction (male of over 50 years of age, smoker, crushing central chest pain) or Parkinson's disease (blank facial expression, shuffling gait, cogwheeling rigidity of the limbs),

and links back to the points raised in relation to the way that medicine is often taught. Some students put significant weight on the demographics, almost over the symptoms, in determining the likely question answer, with comments regarding the highlighting of age and sex as an important part of the process in SBA questions.

Age made a clear difference to students as the patients age would immediately focus them on certain diagnoses being more or less common, particularly things as such as cancer being commoner in the older age group and certain diseases occurring more commonly in younger individuals. Further weight was given to this point, with failure to acknowledge the age leading to judgement errors, and students often said that they would come back to the age after reading the rest of the stem as they might have not processed it effectively. There was some data relating to how this could also lead to some difficulties for students in being able to reason the weight that they should give to the age versus the symptoms themselves, especially where it appeared to be a mismatch with their initial hypothesis.

> "Yeah, you're not sure, you know, like I don't think this should be presenting in someone of your gender, age, or whatever, but like, how much focus should I put on that compared to something else in the question. Like, what's more important, is that more important? I mean, are you telling me it's really probably not that?" *Medical Student, Focus group 4*

The demographics in the OSCE examination had much less impact. Students focussed on the scenario rather than the demographics and stated that they were more focussed on the processing of what the patient/actor was saying rather than their demographic profile. It was also noted that the nature of different OSCEs themselves would throw up challenges that would lead to students actively ignoring these aspects. University 1 medical students noted that in their very large OSCEs (up to 13 simultaneous circuits) the chances of being able to get sufficient "identical" actors/patients was highly unlikely whilst at university 2 they stated that the local area was predominantly white, and more elderly, and consequently so were the actor/patient pool for the examinations. The lack of correlation between the scenario and the actor/patient would lead to cognitive dissonance which led to this reduction in the value and impact of demographics.

> "And if you have like an OSCE patient and they look very slim, but in the history it says your BMI's 40, you discard that because you're looking at them and they look slim" *Student, Focus group 1*

There was also some dissonance noted with specific station types, particularly where the scenario was linked to children. It is generally unfeasible to have children present in an OSCE station due to the difficulty in maintaining engagement, however there is a need to have stations that test knowledge around paediatric conditions. Student stated that this led to challenges of having to engage with a "virtual" child or ignore the fact that they were not there when in reality their presence might lead to different thinking due to the ability to directly observe them and their physical condition.

8.5.2 Ethnicity

Ethnicity was noted to be important. Students stated that ethnicity would immediately trigger different thinking. Whilst age is ubiquitous in SBA questions, and the patient sex extremely common, ethnicity is much less likely to be included. When it was seen within questions students automatically recognised it as crucial in the question and it rapidly narrowed down the diagnostic options.

> "I think I just assume if they're telling me it, they want me to know it. So in the written exam, it was a Nigerian man with pain and I'm like, okay, you want me to pick up on that he's Nigerian, you want me to make an assumption based on that of what's causing his pain. So I feel, I feel like you act differently in exams because you've been given very specific information for a reason whereas if I just saw the same Nigerian man out in the community, yeah, I might have a wider net of suspicion just because it's not an exam." *Medical Student, Focus group 6*

The evidence that ethnicity led to specific inductive assumptions was clear but also impacted differently to other demographic factors, specifically related to the student's own exposure to different cases, and thus is linked to availability bias. One of the SBA questions used in the research was based on a young patient with respiratory symptoms, designed to lead students to bronchiectasis as a possible cause of the patients' symptoms. The students based in university 3 immediately had tuberculosis as their go-to diagnosis based upon their experience within their local area, whereas those from universities 1 and 2 were led to bronchiectasis.

> "I see it the same as pattern recognition, I think, I think that's how pattern recognition works anyway because I'm like I've got this bias because I've, in the past I've seen that these symptoms equal TB and so when I see someone with those symptoms, they have TB until they don't have TB." *Physician Associate Student, Focus group 7*

It is notable that the question did not have an ethnicity recorded within it which also suggests an assumption of ethnicity based on the locality in which they are immersed.

8.5.3 Clinical Setting

The clinical setting of the patient had less impact on the cognitive process than the demographics however it was still linked to the process of initial hypothesis formation.

The assumption of illness severity due to the place the patient presents was shown in the data with students postulating that patients presenting to General Practice would generally be less unwell than those presenting to the Emergency Department. This also manifest itself in the way that students approached the question and the expectations of what was important, and what was achievable in any specific clinical location. Increased weighting of things like vital signs in the more acute setting was expressed, as was the fact that the different settings, with their differing access to certain investigations or treatments, would alter their approach to the decision, essentially altering the expected pattern. It was noted however that the students' own experience in placements was of patients attending either setting with illnesses that were inappropriate with examples such as minor illness in the Emergency Department and an acute myocardial infarction in General Practice. This latter point was cited as a reason as to why the clinical setting was seen as less valuable compared to the demographics.

> "With something like a pneumothorax I'm like, I'll read into the setting that they're in as an idea of the severity of their symptoms. So say somebody's presented with exactly the same history but one time they're in A&E and one time they're in GP, I'm gonna think that their symptoms are more severe if they're in A&E because I'm like this is bad enough that they sought Emergency Care whereas if they're in GP they are well enough to be able to wait to like get a GP appointment and come see me"
> *Physician Associate Student, Focus group 7*

Whilst noted by students, the clinical setting of the question was seen as relatively low in importance in SBA questions. Students stated that they may refer back to it for some aspects of the decision process, especially where two possible answers were to be decided upon, but they often skipped over it as low importance. The clinical setting of the question was seen as very low in relevance in OSCE examinations. Similar to demographics students focussed on the patient in front of them as they recognised the artificial environment of the OSCE itself which mentally detached them from the clinical setting.

8.5.4 Variance with the type of decision

What came through clearly in the data was that demographics are heavily weighted toward diagnostic decisions whilst the clinical setting was more likely to influence management related decisions. Whilst some data linked management-based decisions to demographics, these were around the feeling that less-invasive investigations would be more appropriate in younger patients or certain religious groups. Clinical setting of the question was much more related to decisions around patient management with the recognition that certain management options would not be available in some settings versus others. This also related to the way a question lead in was written. “Most appropriate next step” would lead to the clinical setting being relevant whilst the “definitive” treatment would alter this, and the clinical setting would be ignored.

> “Yeah. Yeah, I think so because like, you would still, you're just, it was just like investigations would be done in a different slightly different order, I think. Um, but then the diagnosis doesn't, is really, isn't really impacted.” *Medical Student, Focus group 5*

Similar to discussions above, the primary relevance was in relation to SBA questions with OSCE not being seen as having the same type of variation and, where questions were asked regarding it, the station itself would lead to a logical conclusion as to what was being asked for rather than it needing the student to refer directly back to the instructions to understand this.

8.5.5 In Test context conclusion

The results demonstrate how the context provided within the question itself impacts upon student clinical reasoning in SBA and OSCE assessments. Overall the impact is higher in SBA questions compared to OSCE questions due to the fact that these are considered to contain all pertinent information and there is no direct engagement with another person. In OSCE stations the fact that the process is more organic and interactive leads to a reduction in the importance of these context factors although some may remain, particularly through either assumption of patient illness severity, or through the decision regarding the most appropriate management outcome.

8.6 Closure Results

Closure refers to the part of the process in which the student makes their final decision and particularly the concepts of satisficing and bounded rationality with the potential biases that these can cause. Its position in the concept framework is shown in figure 14.

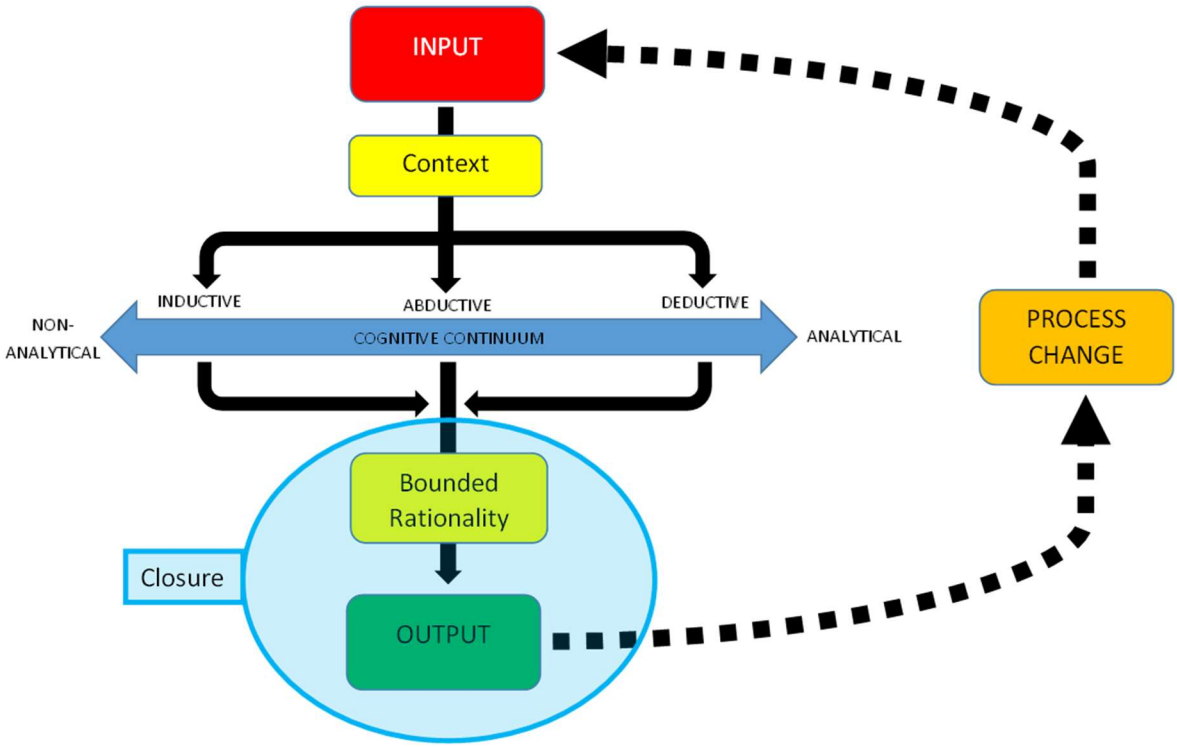


Figure 14: The closure component of the cognitive model

8.6.1 Satisficing and Bounded Rationality

It is worth recapping that the concept of satisficing is that of undertaking a decision-making strategy that aims for a satisfactory or adequate result, rather than the optimal solution, and is the primary tenet of bounded rationality.

Students described that as an assessment progressed, they moved from a position of expecting perfection to one that was more focussed on ensuring a “good enough” answer. This was linked to general cognitive fatigue, and motivation, through the examination but also the fact that as the pressure increased as the exam progressed, and the time to answer questions reduced, good enough was seen as entirely appropriate.

> "Especially when you get towards the end and you are running out of time and you don't have time to read everything properly think (sigh) well, I think at the end of my time whatever C,B,E It looks like it might fit that, yeah so fine." *Student, Focus group 1*

Students noted that the potential impact of the examination on their future career could also have an impact. It is likely that the recent change in the way that foundation places are allocated (Sam *et al.*, 2022) is linked to this comment as prior to the year group studied allocation to the UK foundation programme was via ranking based on course performance but for this group it was based upon preference. There were comments that this shift led to a change in the students feeling that good enough was now acceptable as there was no ranking impact, an external contextual factor directly leading to an outcome related change in cognitive processing. Possibly linked to this it was also notable that some used the satisficing concept as part of their active metacognitive approach to manage their own anxieties. One student commented that striving for the perfect answer across exams would always leave them feeling low at the end of an examination whereas focussing on good enough, with the recognition that just passing was what mattered, allowed them to feel better overall. Another commented that they realised that good enough was fine, particularly in OSCE, as their own personal aspiration was to establish good rapport with the patient and have them feel comfortable and that in doing so it was likely to be reflected in their performance.

The ambiguity of a question, sometimes conflated with difficulty of a question, made a difference. Students stated that perfect was the way they felt they had to be when a question gave them a clear pattern that they could see whereas in questions that were more difficult, or the pattern was less clear, they would accept good enough. Students commented that if they started off well, and were confident, they tended to feel like they were thinking about perfect answers whereas when they came across questions that were clearly more difficult the focus clearly shifted to the concept of good enough.

> "But like I said before again I think it relates to the ambiguity. Like if you don't actually know I think it will be good enough. But if I feel like mmm I've got an inkling it needs to be the perfect answer because I should know this." *Physician Associate Student, Focus group 7*

Some students expressed a very binary view in relation to SBA questions stating that they felt they either knew it or not, with no position that sat between the two. There were a

number of views regarding inductive reasoning where students would feel an answer was right without really knowing why. This was particularly the case when students had been able to eliminate a number of the answers and were considering a smaller subset of answers.

> "There are very few questions where I will have like, just look at all five options and be like it could be any of them. It might be that I kind of am down to two and then I'm like this one just seems more right, and I can't really tell you why, but it just seems more right. But I don't think I would ever do that from like all of the answers." *Physician Associate Student, Focus group 7*

This process was often linked to questions considered more ambiguous or difficult thus linking to the satisficing process. There was also expression that this satisficing process and recognition was linked to the data presented within the question, supporting the abductive reasoning approach.

8.6.2 Differences between exam types

There was a difference in bounded rationality between the two exam types expressed within the focus groups. Much of this was linked to the concepts discussed previously regarding the way that the examinations are constructed.

Students commented that they felt that they had to aim for perfect in the SBA examination as the scoring is binary with either the answer being correct or not. Some commented that an inability to find a perfect answer led to them automatically assuming that they had the answer wrong, with the potential exam anxiety that this would produce. The process for marking in the OSCE however means that many students felt that being good enough was entirely acceptable and that actually getting to a definitive diagnosis or management plan was not crucial to passing the station.

> "So you, in the SBA to get the marks you need to have the perfect answer whereas in the OSCE, as we were saying earlier, I don't think it's necessarily about answers. I think it's about your approach." *Medical Student, Focus group 3*

It is important to note that whilst this was the majority view there were some differing opinions. A small number felt that they felt the need to be perfect in the OSCE examination, however the data suggested this was focussed primarily on their overall performance rather than a specific aspect of decision-making. Several commented that the presence of other parties in the space with them led to a feeling of needing to deliver a perfect performance

to demonstrate their competence. Another felt that the time that they had practised with colleagues to try and deliver an optimal OSCE performance led to their own feeling that they needed to be perfect in the OSCE. For the SBA there was a comment that it was impossible to go for the "perfect" answer because none of them were likely to be perfect (as the concept is single best and so perfection is contextual) and so it was more about trying to find an appropriately good enough choice from the five available. These comments demonstrate that whilst the results show clear tendencies in the two exam types, there are differences seen between students.

8.6.3 Biases

Biases tend to risk poor metacognitive approaches by students. Five primary biases were evident within the data and are defined in table 36.

Table 36: Biases expressed during focus groups

Anchoring Bias	Where we heavily rely on the first piece of information we are given within a question, or a single piece of information we value highest.
Availability Bias	Where we rely on information that comes readily to mind when evaluating situations or making decisions, often based on our own experiences.
Confirmation Bias	Where we choose to ask questions or accept information as confirmation of our existing beliefs or theories rather than seeking incongruent data.
Gamblers Fallacy	This is an incorrect belief that a random event is less or more likely to happen based on the results from a previous series of events
Representativeness Bias	When we make judgments or decisions based on how people or situations match a particular stereotype.

Anchoring bias was predominantly expressed in relation to the OSCE examination rather than the SBA. Students commented that they would latch onto specific words, and it was then difficult to alter their thinking. The link back to abductive reasoning and pattern recognition was also clearly expressed.

> "I think if they give you something super vague to start off with I'm like oh you really want us to dig around here but if they're giving us keywords then I'm like right that's what it is and it's instantly shuttered off." *Physician Associate Student, Focus group 7*

There was clear data showing that the pattern recognition approach and anchoring bias are linked with the potential risk to lead to students to incorrect conclusions due to anchoring to some specific, early, individual pieces of information. This translated into "real world" experiences with comments from medical students regarding pre-conceptions based upon information that students already had from clinical notes regarding frequent reattenders and how this then led to poor decision-making by clinical teams.

Students also commented on their own personal experiences with family and friends own illnesses, as well as themselves, influencing their decision-making due to information immediately being available to them.

> "….everyone has different experiences, different careers before this, and you can bring things from your past experiences into medicine, and just to be aware of this to then take a step back as the common things are common, but that one time it might not be and if you've not really covered everything comprehensively, especially if you're quite junior and then you'd put yourself up for a bit of criticism" *Student, Focus group 1*

It was also notable that local guidance was commented upon as an influence, with one student commenting at their surprise when an antibiotic was given in accordance with local guidance which differed from their previous experience of other prescribing guidelines. Similarly, and as previously discussed, patterns of disease that are more common to a specific local area are linked directly to availability bias and both of these issues have potential implications for national examinations which now occur in both physician associate and medical student training at the point of finals.

Confirmation bias was the most commonly expressed bias that students recognised as occurring and being a potential threat to their performance. There was active expression of ignoring data that was incongruous with the possible answer in SBA questions where students were convinced that they were correct in their initial thinking despite the fact that they recognised that this was metacognitively flawed.

> Confirmation bias is really hard to fight in an MCQ. Like, if you see like two or three things in the stem, and you've seen an answer that fits the things, you know, in the stem, like, I know confirmation bias is there, but I've seen the right answer. Or I've seen an answer I'm pretty certain is the right answer and actually overcoming that in an MCQ and taking your time when there's a degree of time pressure. And you know there's questions where you haven't gotten any, you've got a lot less knowledge. You kind of can't fight that I find. I really struggle to fight it. *Medical Student, Focus group 4*

Whilst much of the response data was in relation to SBA style questions there was also comment on OSCE with observations regarding the challenge it presents in the decision-making process when questions are asked of the patient and the answers do not fit with the expectation, given the hypothesis that has been initially generated.

Students actively acknowledged confirmation bias and expressed that they sought to avoid it and its potential influence, however they noted that for both SBA and OSCE experience of doing practice questions or scenarios also led to difficulties related to pattern recognition. They would see a pattern from a similar question undertaken previously, make the same early hypothesis, and thus follow the same line of thinking that got to the right answer in that case and the risk that entails.

> "If you have like a hypothesis in your mind, I find myself asking questions to affirm, like, what I think it is rather than, like, keeping it broad. But I also find that if you've done like a practice, one beforehand, sometimes you go in and your sort of like oh I've done this recently, I think it could be this one and then you just end up like I'm going down a rabbit hole and asking those sort of questions" *Student, Focus group 1*

Gambler's Fallacy came up in one focus group with two main areas coming out in the conversation, although only one could be considered true Gamblers Fallacy. The first was the concept that there would be unlikely to be multiple questions on smaller topics and the second being answer patterns.

This first observation was expressed by a student noting that they had a question in which they were unhappy with their first answer of sinusitis and therefore when another question came up with sinusitis as the answer they returned and changed their initial question answer. As exams have clear blueprints, it could be seen that this is in fact a heuristic with positive benefit to the student, rather than the gamblers fallacy, as blueprinting should remove the level of randomness that this bias relies on.

The second represents true gamblers fallacy, and it relates to seeing patterns in the responses in SBA questions.

> "Do you find that in an SBA if like four in a row are all A and you're like sure of them then the next one, you're like, you can't be A again, like you've almost kind of ruled it out before you've got to. Actually, I can't, it can't just be A again." *Physician Associate Student, Focus group 2*

It is important to recognise that best practice in writing SBAs for these groups in the UK is that all answers should be in alphabetical order to deliberately randomise the answers and reduce the risk of question writers introducing patterns themselves. As well as repeated answer selections as expressed above there was comment on the idea that if a pattern was observed it might influence the students answer choices and disrupt their thinking.

> Like, if I realise that answers are going in a pattern, I'm like this can't be right, like I must be answering them wrong because they wouldn't put the answers in a pattern. *Physician Associate Student, Focus group 2*

It is clear that there is potential for metacognitive disruption due to this bias and that this could lead to difficulty for students in answering questions effectively.

The final bias was representativeness bias that came through strongly related to pattern recognition and highlighted the potential pitfalls of this cognitive approach. Representativeness bias is a counterpoint to probabilistic reasoning and the general concept that common things are common. The central point with this bias is that just because something is more representative does not necessarily make it more likely.

Students expressed this bias in several ways, including comments on how if they were told the patient was in a poorer area this might influence the range of illness expected to be seen, that farmers did not feel pain in the same way as others so their expression of significant pain should be taken more seriously, and patterns related to ethnicity and weight loss. One of the OSCE stations used in the study was irritable bowel syndrome in a male patient and a student expressed how this affected their thinking as their expectations would be that a patient suffering from IBS would be female, considering this to be more representative.

This bias links directly back to the results on in-test context and pattern recognition as well as aligning closely to both confirmation, availability and anchoring biases in the overall effect the thinking might have on the student's cognitive process.

> "So I think it's that line where you use the bias to your advantage because common is common, hear hooves, think horses as you said earlier but its knowing when you're oversimplifying things and you're relying on that lazy, that lazy process that it must be because it's always this. like it's always this. So I think you have to use it for your advantage and learn how to differentiate between the two" *Student, Focus group 1*

Metacognitively there was recognition of the risks of this bias and the potential to use it to aid thinking whilst needing to remain in control of it to avoid problems occurring from it.

Students demonstrated recognition of these biases and clear attempts to acknowledge and control them in their thinking processes. There was a general recognition of their existence and the dangers that they might present if not managed appropriately. Some of these issues have been previously picked up in relation to the results around preferred cognitive approach. The only other comment is that the results suggest that students use the term bias for any thinking that relates to stereotypes, patterns or the issues raised above, even when the thinking may, in some cases be better described as a heuristic or "rule of thumb" and link to the satisficing process.

8.6.5 Metacognitive development with experience

During the focus groups students commented how their metacognitive control had developed with experience. These related to ensuring that there was recognition of biases and managing them accordingly during the assessment processes. There has already been discussion in this chapter regarding the management of SBA questions with the example of students not spending excessive time on more difficult questions at the expense of losing opportunity on easier ones. Other comments included the need to ensure that there was challenge to the personal thinking process to ensure that stereotypes did not dominate the thinking, a shift from a feeling of the need for perfection to that of good enough, and the external factors that influence and change you.

> "Med school as well you start as a young kid who, for me anyway, has just finished sixth form and I had no idea about the world, but I think like five years is a long time to be studying at med school, I think. In five years you change a lot like getting through sort of like your later teen years into your sort of early twenties and I think, I guess like you said, your priorities sort of change because as you get older your priorities do change like family life changes, and I think that sort of like has an impact." *Medical Student, Focus group 5*

The transition from school to higher education was commented upon by a few students noting that, especially for medicine, you transition from being a high performer at school to being in a group of all high performers and this internal competition can lead to unhealthy behaviours and thinking, which can then affect exam performance. These student specific

context factors clearly influence the metacognitive process with many attending medical, or physician associate, schools during a period of significant change in their lives.

The results overall suggest that as students experience more examinations, and clinical experiences the expectations of themselves, and the approaches they take within the examinations, do undergo some changes directly linked to experience. This is part of the "transition to expertise" that all clinicians will undergo but at its earliest point.

8.7 Summary

This chapter has set out the results of the qualitative analysis utilising the cognitive model developed in the initial narrative review as the structure on which to base the presentation of the results.

The next chapter will bring together the quantitative and qualitative results to help better understand the cognitive model and enable the drawing of conclusions regarding the cognitive approach of these student groups.

Chapter 9 – Convergence and Analysis

9.1 Introduction and Chapter Summary

This chapter seeks to integrate the two sets of data presented in chapters seven and eight and provide a mixed methods analysis and interpretation of the data. The chapter is designed to address the original research aim of obtaining a better understanding of the way students think during assessments and to then allow us to consider the implications for teaching and assessment. Whilst these are new findings, appropriate literature will be used to support the inferences being made. The limited literature available indicates that this study provides new insights.

The chapter will consider the ten questions using the concept framework to provide the basis on which the analysis and interpretation is done to better allow the data to be considered in a coherent manner. The two data sets will be brought together using a similar grouping of concepts as utilised in chapter eight as this provides us with a useful guide through the cognitive process from beginning to end. Figure 15 shows the concept framework with the three primary groupings that will be used to consider the data in this chapter.

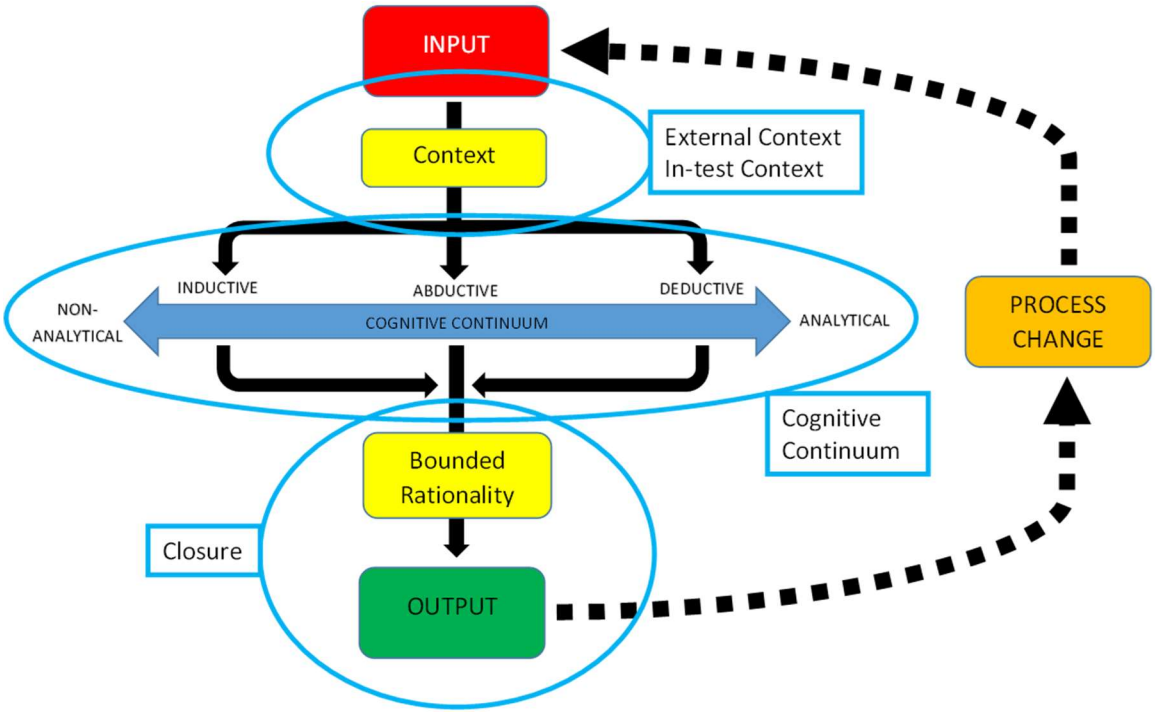


Figure 15: The two data interpretation approaches against the concept framework

Initially Context (questions one to three) will be considered individually. Next the cognitive continuum will be considered (questions four to eight) however there will be a slight difference in data presentation for questions five to eight as these are considered together and the data approached slightly differently as set out in section 9.3. Finally closure (questions nine and ten) will be considered.

This data interpretation process follows the guidance laid out by Creswell and Clark for the interpretation of mixed methods results based upon a convergent approach (Creswell and Clark, 2017). Mixed methods data interpretation in the convergent approach involves the comparison of the two datasets and a process of simultaneous integration (Öhlén, 2011) undertaken to answer two key questions:

1. To what extent do the two datasets (quantitative and qualitative) converge or diverge?
2. Do the qualitative findings materially add to the understanding of the quantitative results?

The use of the same question set for gathering the quantitative and qualitative results facilitates the interpretation process. The data will be presented in relation to these questions, and their underlying element of the reasoning process, utilising a joint display table for each before a narrative consideration of the data, specifically any areas of convergence/divergence. and offering an interpretation. The purpose of the exercise is to develop a comprehensive understanding of the research question.

Following this the discussion chapter (ten) will reconsider the concept framework in light of the findings as well as explore the implications in relation to medical education and assessment relating back to the points considered within the narrative review.

As stated in the methods for the purposes of interpretation and comparison of the quantitative data with the qualitative data the 10 point Likert data scale is considered against the typical descriptors for a five point scale as shown in table 15 (Sullivan and Artino, 2013).

Table 15 – Likert scale descriptors for interpretation

Likert Scale Score	Descriptor term
1 or 2	Never
3 or 4	Rarely
5 or 6	Sometimes
7 or 8	Often
9 or 10	Always

The conversion to these terms allows a better narrative description of the quantitative results facilitating comparison. The qualitative exemplars will have the group from which the text is drawn indicated after the quote e.g. FG1 is Focus group 1.

9.2 Context In the reasoning process

This initial section will consider questions one to three which are concerned with both the external context of being in an examination and the internal contextual factors to consider within the questions themselves. These questions stand alone and so will be considered as such. The joint display tables will show congruent and discrepant data to facilitate the understanding of each individual contextual element.

9.2.1 Question 1 – Student centred context

Quantitative	How I feel in the assessment affects the way I make decisions in questions (Never 1 to Always 10)
Qualitative	How does the way you feel in an assessment affect the way that you make decisions in questions?

The quantitative data suggests that students felt that the way that they felt in an SBA often affected the way that they made decisions. It is notable that the interquartile range is three meaning that the 2nd and 3rd quartiles sat between “sometimes” and “always” thus suggesting there was clear impact on the decision-making process due to being in an examination. In the OSCE the link is stronger with student affect always impacting on their thinking with the narrow interquartile range supporting this.

Whilst the quantitative results suggest that students often felt that their affect in an assessment had an impact on the way that they thought, the qualitative results suggest that this focus was not so much linked to the actual process of clinical reasoning but more that being in an assessment led to greater uncertainty for many. The phrase “second-guessing”

came up multiple times in the data pertaining to this area with the implication that the position of being in an assessment led to greater anxiety as to whether any specific answer was likely to be correct. Confidence in the topic of a question, or the way that an OSCE station flowed, was clearly linked to cognitive ease and the impact of this on the individual's confidence in the answer or the process.

OSCE was often described as a performance, and it may be that personal character has more of an impact in this type of examination (Shin *et al.*, 2011).

Table 37: Question one joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 7 Mode – 7 Mean – 7.36 IQR – 3 **OSCE** Median – 8 Mode – 8 Mean – 7.59 IQR - 2	Uh, yeah, I, I would probably, I think I probably agree (FG3) I think things are hard actually knowing your careers depending on that exam (FG6) For me if I feel good before an exam, or doing a question, I'll be confident in my answer, but if I'm not so sure, if I don't feel as prepared then I'll second guess myself a lot when I'm answering a question (FG7) It does because in the real world you're more relaxed. You take your time if and if you take like 12 minutes, nobody would care (FG2) When you're revising for an exam it's always a textbook definition of this, a textbook definition of that, but when you're on the ward it's, nothing's ever textbook. So like, and I think exam questions are always focused on textbook definitions (FG5)	…the OSCE is all about confidence. I, what I feel like it's all about confidence. You could, even if you don't get everything, I think your confidence says a lot about you as a person. (FG1) …you've just got a lot more pressures on yourself, that's a bit more adrenaline. So I think sometimes that helps with making decisions because it feels like you're very zoned in… (FG7) So as long as I'm safe in the seven minutes and try and get as much done without seeming unsafe or inappropriate then, you know, that's as much as i can do. (FG3)

Students commented on the feeling of needing to do well for the examiner and the patient, highlighting that the presence of another party in the examination may contribute to the

difference seen in the quantitative results. This is not dissimilar to the work of Skellern (2020) and the link between emotive content and biases in reasoning.

Recent work on situated cognition (Penner *et al.*, 2024) discusses the fact that context forms one of the key parts of the decision-making process with time pressures specifically cited as reducing cognitive performance, with evidence of this leading to a divergence from the effective use of clinical guidance (Tsiga *et al.*, 2013). This gives rise to a reflection as to whether the time pressure applied in these examinations adversely affects the process of clinical reasoning in such a way as to lead to some of the behaviours, and process preferences, that have been articulated through the focus groups.

The order of the questions, particularly of SBA questions, contributes to the pressure with students commenting that hard questions would lead to increasing anxiety and a detrimental effect on cognitive performance. This has previously been seen in first year medical students in a study looking at implications on exam performance (Pradhan *et al.*, 2014). With the increase in computer-based exams, and the ability to randomly order the questions, this has clear implications as some students may be detrimentally affected by the randomisation if particularly difficult questions start the assessment. This links to the work of Durning et al (2012b) with the altering of specific contextual factors leading to differing effects on diagnostic reasoning. It is difficult to see how this could be avoided in an OSCE where the nature of it being a circuit will inevitably lead to different students starting on different stations with some being more, or less, to their liking with the resultant variation in stress.

It is important to note that some of the students expressed a more pragmatic view, particularly with the OSCE examination, where there were comments linked to the recognition that the mark scheme in any one station was cumulative and thus it was about performing as well as possible within the time available. This approach, focussed on maximising marks through good performance, has the potential to reduce the stress felt, however it can lead to points-scoring behaviours amongst participants that may move away from a patient focussed interaction (Gormley *et al.*, 2021). Similarly in the SBA examination some described strategies to target maximising scoring by answering all the questions they were confident with first, then returning to the start and answering the harder ones. In doing so they were they were seeking to manage their cognitive ease.

The qualitative data supports the quantitative data however, as stated, it may be that suggesting that there is a true change in cognitive process is incorrect and it is better described as the students being moved from a point of eustress (optimum performance) to distress in which their performance drops (Cassady and Johnson, 2002).

9.2.2 Question 2 – Patient centred context (demographics)

Quantitative	The patient demographics have an impact on my decision-making (Never 1 to Always 10)
Qualitative	How do the patient demographics have an impact on your decision-making

The quantitative data suggests that patient demographics in examinations often contribute to clinical reasoning in both SBA and OSCE. The narrow interquartile range and mode, which is in the "always" range of the Likert data item, suggests this is a strong contributor to clinical reasoning in the SBA assessment, whilst less so in the OSCE assessment, with its lower mode and wider interquartile range. The qualitative data supports this difference between the two exam types with the demographics being much more strongly related to student clinical reasoning in the SBA examination when compared to the OSCE.

Arguably the qualitative data suggests more of a disparity between the two results than the quantitative results would suggest, and this is highlighted in table 38 where the congruent data is SBA focussed whilst the discrepant data is OSCE focussed.

In SBA assessments students recognise that question writers put information in the question stems that are relevant, and avoid things that are not, in order to ensure that the question stems are not excessive in length and to focus on pertinent information. Combined with the recognition that SBA questions are usually focussed on "textbook" illness scripts this naturally leads to pattern recognition inductive clinical reasoning being prevalent amongst students in this type of examination. Ethnicity is a good example of this. Students recognised that the presence of an ethnicity in a single best answer question meant this was going to be linked to the answer directly. A previous study from the USA specifically advised that ethnicity should only be used on medical licensing examinations where it has direct relevance to the individuals health (and its social situated context) reinforcing this situation (Ripp and Braun, 2017).

Table 38: Question two joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 8 Mode – 10 Mean – 8 IQR - 2 **OSCE** Median – 8 Mode – 8 Mean – 7.26 IQR - 3	I think it immediately, whether consciously or not, it immediately like focuses your mind on certain differentials. (FG3) And I think sometimes I put a lot of weight on the demographics rather than the symptoms. So I've kind of made my decision before even looking at what symptoms they even have. (FG2) I know there was an SAQ question where I saw the patient demographics and I instantly, without reading the rest of it, I kind of thought I knew what the….yeah (FG4) I think I just assume if they're telling me it they want me to know it so in the written exam, it was a Nigerian man with pain and I'm like, okay, you want me to pick up on that he's Nigerian, you want me to make an assumption based on that of what's causing his pain. (FG7)	I think in an OSCE I'm so busy thinking about what I'm going to do next. I just don't even think about their age, not at all (FG2) It does make a difference to come to the diagnosis. And when you're actually reading the SBA, you do consider it, but I find in the OSCE when you go and you introduce yourself and you ask them for date of birth or age. And then while taking the information, I'll sometimes don't consider it. I tend to forget. (FG1) Less so in an OSCE, but a, yeah, probably just more about rapport and trying to get answers and the way I ask my questions changes but I don't think it directly affects the way I'm thinking (FG6) you can't have like 15 identical actors for the circuits they're all going to be slightly different ages, slightly different levels of wear and tear visible on them, possibly different sexes, possibly different ethnicities and so you have to assume that some of the factors that you're picking up on in the person, you're seeing are not standardised and if they're not standardised, then they can't significant. (FG7)

Similarly the risk exists of creating incorrect ethnicity-based disease associations and maintaining cultural stereotypes with inherent risks of harm (Cerdeña *et al.*, 2022). With the current focus on the diversifying of the medical curriculum there is an inherent challenge in balancing the need to assess across the range of clinical care, including health issues that are directly related to, or at higher risk in, certain ethnic groups and the issues it can present

in SBA assessments where it can drive pattern recognition and risks of representativeness bias.

OSCE examinations are different due to being interactive and the students having to gather information from a simulated patient as part of the process. Students reported that they would look at the demographics in the details of the OSCE station prior to entering the station as part of a process of formulating hypotheses to test in the station however there was often a discrepancy between the information provided and the actor/patient in the station. Age, habitus, and ethnicity were all mentioned as areas of mismatch between the station instructions and the patient or actor in the station. This led to comments that once in the station the impact of demographics was significantly reduced by this lack of standardisation. This was taken further with students in one focus group (university 2) noting that the area in which their school was based was majority white and elderly and thus the patients in the OSCE examination invariably fitted to this type. Along with the challenge of finding standardised patients across multiple circuits these issues clarify the difference, and potentially challenge the similarity of the quantitative results, for the two assessment types studied.

Regarding patient demographics there is good correlation between the quantitative and qualitative results for SBA but less so for OSCE examinations where the quantitative results seem somewhat high compared to the narrative gathered from the focus group participants. In both cases the demographics form a key part of the cognitive process with respect to the formation of initial hypotheses around likely diagnosis but in the SBA, where there is no information gathering component as such, this is much stronger than in the OSCE. This would fit with the nature of the two assessment types with OSCE being more evolving and, increasingly, having multiple facets (e.g. focussed history and examination) versus SBA questions which test a single element of knowledge.

9.2.3 Question 3 – Patient centred context (clinical setting)

Quantitative	The clinical setting of the question has an impact on my decision-making (Never 1 to Always 10)
Qualitative	How does the clinical setting of the question have an impact on your decision-making

Similar to demographics the quantitative results suggest that students often consider the clinical setting in the question though the slightly lower mean and wider interquartile range suggests that again this is stronger for the SBA than the OSCE examinations.

Table 39: Question three joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 8 Mode – 8 Mean – 7.6 IQR - 2 **OSCE** Median – 8 Mode – 8 Mean – 7.27 IQR - 3	If you're thinking A&E and somebody comes with abdo pain my head is automatically mostly thinking something more serious than somebody who comes into GP (FG4) So say somebody's presented with exactly the same history but one time they're in A&E and one time they're in GP, I'm gonna think that their symptoms are more severe if they're in A&E because I'm like this is bad enough that they sought Emergency Care (FG7) Yeah. I guess the only difference would be if they ask you like you know the questions at the end of there like what would your management be? Yeah. Like i thought that you have to have in the back of your mind like where you are because if they go, oh what would your management be and if it's someone who you think is having a heart attack then you need to refer to ED (FG2)	At the end of the day, it's not something that affects how I answer or at least not like consciously and then even when I was doing questions it would tell me the setting but I think I just assumed that people don't know how their health care system works then apply that to everything. (FG6) I suppose in a more SBA if it's like, maybe I'd pay a little bit more attention, but a lot of the time its if it's this hospital ward and A&E they're gonna be slightly different things that are going to be looking for, but I wouldn't say I ever took that much attention to it in the questions (FG4) I think with OSCEs as well you sort of, once you've done a few stations you completely forget where you are ……I think, yeah, like you said you're just in a blue box really. (FG6)

The qualitative results do not comfortably support the quantitative results with many comments stating that the clinical setting contributed less strongly to the clinical reasoning process than the statistical results suggest. The setting was seen as just something that

formed part of the stem, some describing it as white noise, which is at odds with statements made in response to other questions where students stated that everything in the stem was relevant.

Where it was considered, it was SBA in which it was expressed most clearly, with students stating that they made assumptions of severity based upon the clinical setting of the question as the Emergency Department was seen as higher acuity than a General Practice setting, and therefore patients were expected to be less well. In OSCE students often expressed that the "setting" was that of being in an OSCE (e.g. a blue box as stated in table 39) and the setting itself became an irrelevance with respect to undertaking the assessment process. Even where the setting was noted the impact on the clinical reasoning process was extremely limited with students stating that the setting itself did not have any real impact and one noting that in the current healthcare system patients frequently attend the "wrong" place, particularly lower acuity patients attending the Emergency Department.

It was clear that the impact of the clinical setting was mainly focussed on management decisions where the choice of investigations or specific drug treatments could be directly influenced by the setting in which the patient is seen. An example given was cluster headache as in General Practice a triptan would likely be used, whilst in the Emergency Department short burst oxygen therapy might be preferred (Peng and Burish, 2023). The availability of investigations was also seen as differing and thus the treatment plan might vary. This has interesting implications for OSCE stations set in the Emergency Department as over-investigation in this setting is the subject of much literature and leads to concerns around diagnostic stewardship for which high quality reasoning is crucial (Fabre *et al.*, 2023). It suggests that including the setting, with its triggering of expectations that the patient is more acutely unwell, might lead to students making assumptions regarding the needs for more, or more complex, investigations than are appropriate and the risk that the assessment process leads to a self-fulfilling prophecy of reinforcement of this high investigation approach with potential for patient harm.

Regarding the clinical setting, the combined results suggest that the quantitative results are possibly somewhat high when considered against the qualitative data. The clinical setting of the question appears to have greater impact in SBA questions than in OSCE stations and are primarily of value when considering the management of a patient rather than the diagnosis.

The uniformity of the OSCE setting will also contribute to the lack of impact that the clinical setting has in this assessment type and, where included in the station information, it appears not to have much impact in the process due to the task orientated focus of the exam itself.

9.3 Cognitive Continuum

The cognitive continuum considers questions four to eight which all pertain to the continuum of reasoning from purely deductive to purely inductive. This section will also consider early hypothesis formation as part of this as it is strongly linked to inductive reasoning. With inductive and deductive reasoning being each other's congruent and incongruent data the joint display table for questions five to seven will be presented as a quantitative data table and separate qualitative display table with inductive and deductive rather than congruent and incongruent data. Questions four and eight will be presented individually as they stand more alone.

9.3.1 Question 4 – Early Hypothesis Generation

Quantitative	I usually come up with hypotheses immediately then change them as the question develops (Never 1 to Always 10)
Qualitative	Do you usually come up with hypotheses immediately then change them as the question develops

The quantitative results suggest that students sometimes, or often, generate early hypotheses when answering exam questions with the data suggesting that this is slightly stronger for SBA than OSCE with the higher mode, mean and the narrower interquartile range.

The qualitative results suggest that student's interpret early hypothesis formation as pattern recognition which is likely to have led to the higher score for the SBA questions. Comments regarding having "seen this before" when doing revision questions from banks such as PassMed reinforce this view and, it can be argued, that this is entirely appropriate as in a clinical setting this is a part of the "expert" thinking linked to inductive reasoning. The concept of early hypothesis generation is often triggered by a very small amount of information provided at the commencement of any encounter, for example an Emergency Department triage note, or the opening presenting complaint articulated by the patient. The

qualitative data from previous questions supports that students see SBA as primarily a pattern recognition process and the nature of short stems leads to limited opportunity for early hypothesis formation as the question itself is the only information provided for the student to consider and thus the clinical reasoning involved in the question itself is the only opportunity for hypothesis formation to occur. Students talked about rapidly coming to an answer in SBAs, which is in line with early hypothesis formation, and then being comfortable if the answer was present in the available options.

Table 40: Question four joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 7 Mode – 8 Mean – 6.86 IQR - 2 **OSCE** Median – 7 Mode – 7 Mean – 6.52 IQR - 3	In regard to OSCEs I do like to go in with a few differentials first and then I either rule things in or out. It helps me structure the questions I'm going to ask (FG1) I think it's just characteristic of the medical profession like there are things that you become aware of as you go through that unfortunately you can't get away from and…. It does, it plays a part in what you, like you're thinking, like your mindset as you go in to see somebody. (FG6) I feel like in a SBA if you read a question. If, if you, if something pops out to you that's most likely the right answer. It's before looking at the answers because you've read the question and something's gone in your head and be like, it could be this whereas then you look at the answers and then that sort of either confirms it or, err….points you to another direction. (FG3) I think, yeah, I think I personally do as soon as I start reading I'm starting to think what could be going on with this? (FG5)	In the SBA where you're kind of given all of the information you need to answer the question you don't necessarily immediately have to have an idea because, you know, you've got all the information in front of you (FG1) I didn't really come up with one or the other first I was, just that was one of the times where I was, I finished the question and then started thinking so I think it depends on the question and, I guess, your knowledge. (FG4) I think it'll be slightly different as well when we have real patients like we will in our final exams, but in most of the ones that we've done so far there's not been any real pathology so probably when you're seeing the pathology you're then going to recognise, okay this person is clearly showing the signs of x condition, but at the moment when we're all just kind of being our own patient or we've got like actors who don't actually have any pathology, it feels harder to do any kind of pattern recognition. (FG7)

OSCE is different to this. The qualitative data demonstrates that students take the initial information and generate early hypotheses in order to guide their thinking and question formation with the feeling that the station could "go anywhere". This is much more in line with a normal clinical encounter and links to the concept of abductive reasoning with early hypothesis formation from an initial amount of data determining the point of entry to the cognitive continuum and initial thinking more toward the inductive or deductive ends of it depending upon this data. Clearly this process will also depend upon the type of OSCE station with history and diagnostic communication stations likely to be the ones that would relate most closely to this. There was also data indicating that students were left with a feeling of discomfort in OSCE situations where they found themselves unable to effectively hypothesise from the provided information and therefore were unable to enter the station with a place to start.

It is notable that students had a clear awareness of the potential link between early hypothesis generation and premature closure and the risks this could produce. This will be discussed more under question ten.

Considering the two data types therefore we must conclude that early hypothesis generation occurs, however in SBA questions it is the primary process being undertaken as the student reads the stem whereas in an OSCE station there is early hypothesis formation based upon the details provided outside the station which then guides questioning within the station, supporting the student in ordering their own thoughts appropriately and thus guiding them in their approach.

9.3.2 Questions 5,6 and 7 – Deductive and Inductive Reasoning

Quantitative	Q5 I usually generate a list of possible answers and eliminate them one by one Q6 I try to keep an open mind until I have formulated an answer Q7 I look for recognisable patterns within the question to generate my answers (Never 1 to Always 10)
Qualitative	Q5 Do you usually generate a list of possible answers and eliminate them one by one? Q6 Do you try to keep an open mind until I have formulated an answer Q7 Do you look for recognisable patterns within the question to generate my answers

The quantitative results suggest that students sometimes/often generate a list of possible answers and then eliminate them one by one and similar results are seen for students seeking to keep an open mind.

The results for keeping an open mind are stronger for OSCE than they are for SBA, suggesting a tendency toward a more deductive reasoning approach within these examinations. The higher mode and narrower interquartile range for question six applied to OSCE reinforce this. The results on question seven suggest students often use pattern recognition (inductive reasoning) in both types of examination although the slightly higher mean would suggest that this is slightly stronger as a reasoning approach in the SBA examination.

Table 41: Questions 5,6 & 7 Quantitative data

	SBA	OSCE
Question 5	Median – 7 Mode – 7 Mean – 6.54 IQR - 3	Median – 7 Mode – 8 Mean – 6.49 IQR - 3
Question 6	Median – 7 Mode – 7 Mean – 6.49 IQR - 3	Median – 7 Mode – 8 Mean – 7.18 IQR - 2
Question 7	Median – 8 Mode – 8 Mean – 8.30 IQR - 2	Median – 8 Mode – 8 Mean – 7.98 IQR - 2

The results for keeping an open mind are stronger for OSCE than they are for SBA, suggesting a tendency toward a more deductive clinical reasoning approach within these examinations. The higher mode and narrower interquartile range in question six for OSCE suggests that there is a stronger tendency toward keeping an open mind in these examinations when compared to SBA. The results on question seven suggest students often use pattern recognition (inductive reasoning) in both types of examination although the slightly higher mean would suggest that this is slightly stronger as a clinical reasoning approach in the SBA examination.

The qualitative results support the quantitative results in the main, however it is important to note that the difference between SBA and OSCE appears more pronounced once the qualitative data is considered in more depth.

Table 42: Questions 5,6 & 7 Qualitative data joint display table

Qualitative Deductive Reasoning	Qualitative Inductive reasoning
So I'm the same so I'll read and I'll kind of read sentence by sentence and then say, there's the first bit of information, right? What could this be, this is likely to be ABC, and I'll read the next set of symptoms and it's like alright this could be A or B because it doesn't fit with C. And I'll go back through, and I'll say right, I've got these two or three things in my head. Then I'll look at the answers and then if one's there great. If two's there, but I'll go back to try and see if there's any like information which could tear it apart, you know like age, like sex like race you know? (FG1) with the OSCE If you figure out what it is then you'll forget all the other questions, whereas the best OSCE stations are the ones you don't know what the condition is because then you literally rule out every single thing which is what the OSCE examiner wants to see, that you're asking the right questions. And then after you've asked all the questions and your like, oh, it could be this. (FG3) I can't be so confident to just be like, yeah that's it. I have to go through every answer to like rule out every single one (FG2) In an OSCE setting, I think I definitely… and if like somebody came in with like, I don't know, a headache or like chest pains or whatever then, you know, I'd have like some ideas in mind but I would, I wouldn't form like a hypothesis really early. It would be like much later on. As I got more information I would paint a picture and then be like this fits most with this, this fits most with that. (FG5)	I look for patterns in everything to generate my answers. Like I feel like a lot of medicine is just patterns. (FG6) I find that when I go into an OSCE situation that I have a very structured plan about where I'm going and how I'm asking questions. Then I find if the patient kind of goes off on a tangent I lose a lot of confidence and it's just going back and then having to go back and then I find it can then unravel my whole examination (FG1) I think yeah definitely. I think a lot of…when you do like on Passmed it is a lot of pattern recognition and I think sometimes you don't even read the stem. You see like a buzzword like, I don't know, Reed-Sternberg cells, and as soon as you see that you know that's like Hodgkin's lymphoma. (FG5) I think medicine is pattern recognition really isn't it and that's why the more experienced you get the more a consultant will say oh yeah I've seen this however many times before so I know that this is, we should do this or whatever and because I think it's the more you see it, the more things come up or the numbers come up or a patient presents in a certain way. (FG4) I would still say in quite a few questions I want the perfect answer and particularly the ones where they have that set pattern that you can really recognise you can look at it and go ok it's like everything that you're expecting for gallstones and the pain matches, the like length that matches, the way they're describing it matches, like it all fits. (FG7)

The structure of SBA driving pattern recognition, along with the way students learn, features strongly in the qualitative output when compared to OSCE. The recognition that all the information is provided within the stem, and the fact that the guidance is to avoid unnecessary information, drives a strong pattern recognition (inductive) approach to clinical reasoning in these assessments. The use of question banks reinforces this, especially when it is considered that there are only so many ways to write clinical questions, which must ensure appropriate evidence base, and thus there is inevitable repetition.

Whilst OSCE also shows this tendency toward pattern recognition the qualitative data is more nuanced with the data around patterns being more about structuring the station and seeking patterns through exploration with the patient and seeking specific key words that could suggest a specific diagnosis. The nature of the OSCE is also such that the history can be deliberately less clear cut, patients need only volunteer information when asked, and they may more closely relate to a true clinical consultation in being multifaceted and less "single task" orientated than in SBA.

When we consider the answers to these three questions together, we see clinical reasoning processes that are inductive, but we need to recognise that we have evidence of students managing their thinking behaviours - metacognition. Some are actively trying to reduce their impulse to look for a pattern by deliberately employing a rule out process when undertaking SBAs and there are comments regarding the impact of knowledge on the process.

In OSCE we see comments about the risk of missing things if pattern recognition drives reasoning (linked to premature closure) and thus we can see the nuances that lead us to conclude that the reasoning process must be on a continuum with the individual and task factors determining where it is entered.

If we consider the results of these questions in relation to the concept of abductive reasoning, then these conclusions make sense. Abductive reasoning is about using the data available to reach a most likely conclusion and is not truly a midpoint between induction and deduction but rather takes elements of both. Abduction contains elements of perception (I notice something may be the case), imagination (I am able to realise that this could be a specific case pattern) and creativity (I am able to create a working hypothesis from this information) (Veen, 2021). It is this latter part of the abductive process that then

will contain active elements of induction and deduction with differing amounts of either. Returning to the point made in the narrative review, that dual process theory is illustrative with no true reality, abduction provides a better explanation of the clinical reasoning process. Accepting that an SBA contains all the data that there is on which to base the decision means students will notice key features, be able to imagine how these link to their stored knowledge and then use this to create a coherent explanation – in this case via an inductive process. In an OSCE the same is occurring more iteratively and thus students are driven to a slightly more deductive and creative approach. Students are observing information as it comes in and imagining how it fits into possible diagnostic patterns to create their hypotheses. In both cases this process is entirely information, and student knowledge dependent.

This helps us to accept why indicators of both deductive and inductive reasoning score relatively similarly but with pattern recognition slightly stronger in both assessment types. The reality is that a focus on the two concepts as separate thinking approaches is too simplistic and thus prone to confusion in our own minds.

The data suggests that the clinical reasoning process is best described as abductive with induction and deduction simply approaches within the abductive method. Recognising this allows us to simplify the explanation of clinical reasoning and provides a unifying concept across assessments as will be discussed in more detail in chapter ten.

9.3.3 Question 8 – Inductive Reasoning/Pattern recognition

Quantitative	I tend to come an answer but am not sure how I got there. (Never 1 to Always 10)
Qualitative	Do you tend to come to an answer but find that you are not sure how you got there.

Whilst related to the previous two questions this question is being considered separately due to the significant difference in the quantitative results when compared to questions six and seven. In this case the responses for both SBA and OSCE showed that students rarely/sometimes found that they came to an answer without truly understanding why, a significant difference from the previous questions despite its link to inductive reasoning.

The qualitative data shows that this occurs more in the SBA examination than the OSCE, which correlates with the quantitative findings, where the mean and median for SBA is slightly higher than for OSCE. Students expressed that the nature of OSCE as an information gathering process naturally led to them not having to make these more "gut feeling" decisions as they had followed a process to get to the point that they were at. Building on discussion of abduction in the previous section this "gut feeling" is linked to the imagination component of the abductive process which then drives the creation of our patient management. Interestingly the concept of "gut feeling" and its impact within primary care has previously been explored with a conclusion that it had a significant role as part of the diagnostic process in the creation of feelings of both reassurance and concern (Stolper *et al.*, 2009).

Table 43: Question eight joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 5 Mode – 4 Mean – 4.83 IQR - 3 **OSCE** Median – 4 Mode – 4 Mean – 4.53 IQR - 3	I think sometimes when you do in an SBA. When you do read a scenario there were, there were a few times where I'd read it and I'd think, oh my god, I have no idea what what's going on. I'm not sure what they're trying to ask me. And then you'd look at the answers and you'd immediately rule out three (FG3) Yeah. It'll just feel like a word has triggered something in your brain and it's like, ah, that's the answer but I have no idea why that word has popped up in my brain but it's patterns again isn't it. (FG6) I get to a lot of answers and I'm not sure, quite sure how I got there (*Okay*). No and I honestly sometimes I feel like it's just right. You know how you said vibes (Yeah). Um, sometimes like I'll just look at a question and I'll just be like, well, the vibe that it's giving off I feel like it's an MRI spine, right? (FG7)	It doesn't happen very often just occasionally, I'm like, I'm like there's something in the back of my head and like something here is connecting to this. (FG2) Like in medication questions, like, what's the first line medication then I might like recognise it but I don't know why but in like normal questions I think no (FG2) I'm like very rarely surprised by like where an OSCE stations actually gone like cos I feel like I've done the process of doing the history. I've done the exam so it's kind of…they might have said something that I'm like oh that's new, but I kind of led it there so it kind of feels more organic (FG4)

In this respect the fact that students felt this less in OSCE may suggest that our assumptions about OSCE providing a better window on practice through a "shows how" approach may be flawed when compared to the reality of practice. This is perhaps inevitable given the use of actors, and requirements for standardisation across multiple circuits in an OSCE setting. The data pertaining to the SBA was slightly different with evidence that in some questions there was often a feeling of choosing an answer because it felt "more right" than another although in these cases it was usually in relation to having eliminated several prior to making that choice. Students expressed that sometimes it might be something that they had heard but not looked at specifically within their revision, examples included drug names or specific management approaches, and in the examination it just came back to mind.

There was some comment that this occurred more as students came under increasing pressure in the assessment, particularly time pressure. In this way this phenomenon might be considered as part of satisficing rather than inductive reasoning. The student is choosing an answer that feels "good enough", usually because they do not have the required knowledge immediately available, and thus this is a satisficing process rather than an inductive one. Comparison was made to the clinical environment where sometimes you might see a patient and, although their observations are satisfactory, something just does not feel right to the clinician and so a course of action is chosen accordingly.

Concluding this question the quantitative and qualitative data do seem to align however the question output is more aligned to the satisficing process (to be discussed more in the next question) than the cognitive continuum.

9.4 Closure

Closure considers questions nine and ten which pertain to the end of the clinical reasoning process in which the information is brought together in formulating the final decision. It is notable that, as commented above, question eight might be thought of as pertaining to this area as well.

9.4.1 Question 9 – Satisficing/Bounded Rationality

Quantitative	I aim for a good-enough answer rather than a perfect one (Never 1 to Always 10)
Qualitative	Do you tend to aim for a "good-enough" answer rather than a perfect one?

The quantitative results suggest that students sometimes went for a good enough answer rather than a perfect one. This was seen for both SBA and OSCE however it is notable that the mode for OSCE was higher than for SBA (more students selected 8, in the "often" range, than any other) but the interquartile range was the widest of any question in this study suggesting significant variation between students.

Table 44: Question 9 Joint Display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 6 Mode – 7 Mean – 5.86 IQR – 2.5 **OSCE** Median – 6 Mode – 8 Mean – 5.57 IQR - 5	Think it helps my expectations, if I don't strive to get the perfect answer in all of them because the exams are quite hard, often harder than the practice questions I've done so if I go in and expect to get all of them right then I'll come out and just be really sad (FG1) I think in the beginning I was like I really want to do really really well and like get it all perfect and then that quickly went away. Um and, um and then I was like actually what really matters to me, like, so the good enough is like oh good enough, I'll pass the station type of thing and like that's fine (FG5) I think my approach to the OSCE was, well I just need to pass, don't do anything stupid and it should be… everybody says that not many people fail and it should be fine so just don't do anything stupid and you'll be fine. (FG4) i think OSCEs, the way that I told myself is OSCEs are there to make sure you're safe as an F1. And as an F1 it's not the most crucial piece of thing that you make the diagnosis, or you don't. It's sort of can you sort of pick up the red flags or the major signs, and can you make a decent management plan, and can you have some sort of idea what's going on. (FG3)	I think the reason why I run out of time or have lots of questions to answer in the last five minutes, is because I'm trying for it to be perfect in the first half and then I'm going to the other end of extreme ok like I can't keep doing this I haven't got enough time. I think it's really hard not to be perfect because you want to make sure you get every single mark. (FG1) In OSCE I go for perfect (FG6) in an SBA, you can get, if you get the diagnosis wrong, you've probably got the answer wrong (FG7) I think if I'm doing an SBA it's always good enough is fine. Because it's just me and there's like hundreds of questions but in OSCE it's got to be perfect because there's like a patient there (FG7) I think it relates to the ambiguity. Like if you don't actually know I think it will be good enough. But if I feel like mmm I've got an inkling it needs to be the perfect answer because I should know this. (FG2)

The qualitative results help us to understand this in more detail. The data suggests that in SBA students see the questions as binary, you either get them right or wrong, and therefore there is a general perception of the need for perfection as anything else is not going to score them any marks. The OSCE is a different assessment environment in which students are constructing their diagnoses or hypotheses and, in this respect, have more in common with the clinical environment. In OSCE students recognise that it is about achieving satisfactory performance across a range of scoring opportunities within the station and, due to this, perfection is not required. This is a good example of the satisficing process. Conversely the nature of the OSCE, with the presence of a patient and examiner led to some feeling that they need to be perfect due to the external presence in the room, perhaps indicating more of a performance anxiety rather than anything linked to the reasoning process.

Revisiting SBA there is clearly a shift in acceptance of "good enough" in two key elements, namely when time pressure becomes significant and when the knowledge is not there to be certain on the answer to any particular question. As students became either uncertain, or they had multiple questions left to answer in a short period of time, they became happier to accept "good enough". In this respect we are seeing the impact of the assessment context where the satisficing process itself is impacted by situational elements related to the examination.

We can conclude that the assessment context, and shifting elements within it (question complexity, time pressure, exam type) will directly impact the satisficing process. In this respect the original 1955 construct of Herbert Simon holds true, namely that we have a number of variables, we ascribe value to these variables and within the time available to us make a decision most suited to fulfil the requirements of the situation (Simon, 1955). In this respect the answer to whether students accept "good enough" is yes, sometimes, but there are assessment specific elements that will directly influence this.

9.4.2 Question 10 – Biases

Quantitative	I acknowledge and try to avoid cognitive biases in my thinking (Never 1 to Always 10)
Qualitative	Do you acknowledge and try to avoid cognitive biases in your thinking?

The quantitative results suggest that students often recognise and seek to avoid cognitive biases in both SBA and OSCE examinations. The results are extremely consistent between the two examination types.

Table 45: Question 10 joint display table

Quantitative Result	Qualitative congruent data	Qualitative discrepant data
SBA Median – 7 Mode – 7 Mean – 6.42 IQR – 3 **OSCE** Median – 7 Mode – 7 Mean – 6.78 IQR - 3	I think once you see it then it's in your head, you just focus on that one thing you won't go, oh, there was something else there. (FG7) I think talking about this area I think it's dangerous when you fail to acknowledge other signs. Say, it could be something else. And you so just go down that road of this is what I thought it was. This is the first thing that came into my head, so it is this, I think it's hard not to sort of ignore that, that voice that sort of pops in when you see that question, it really is. But I think if you're not challenging it that's where the problem lies. (FG1) I think it's that line where you use the bias to your advantage because common is common, hear hooves, think horses as you said earlier but its knowing when you're oversimplifying things and you're relying on that lazy, that lazy process (FG1) it's just like factors that you have to be aware of like you live in a poor area and you're more likely to get these conditions, but that doesn't mean that you're only going to get these conditions, and you do have to consider other things (FG6) Going back to the basics of sort of like history taking in OSCEs. I always try to avoid like leading questions because you can narrow down all of your options incredibly quickly (FG3)	I think if they give you something super vague to start off with I'm like oh you really want us to dig around here but if they're giving us keywords then I'm like right that's what it is and it's instantly shuttered off. (FG7) I think once you see it then its in your head, you just focus on that one thing you won't go, oh, there was something else there. (FG2) You know when you're kind of convinced it something and you're like Oh you kind of ignore information because you're just wanted to be this answer. (FG4) I mean, my mom always says stereotypes exist for a reason, like they've come around for a reason. But it's hard not to use bias. (FG2) I know you can get to the point where you form one differential in this sort of a biased way you think of a differential and you pick out the symptoms to match that diagnosis that you're thinking of. (FG4) It's like a massive like thing in the back of my head, like if it's like ABCD I'm like well the next one can't be E that would be ridiculous. (FG2)

The qualitative results confirm that students recognise that cognitive biases exist, however in much of the focus group output the comments focus upon the difficulty in resisting many of the cognitive biases that occur, rather than directly commenting upon any action seeking to prevent these biases impacting on the reasoning process. In this respect representing this data in the joint display table was challenging as the congruent data illustrates student's acknowledgement of bias but it is difficult to ascribe these statements to direct avoidance action as a metacognitive process. As described in the qualitative results five main biases were recognised by students but often with specific reference to the difficulty in avoiding these biases rather than indicating that actively seeking to avoid biases is part of their cognitive process.

It would be appropriate to say that gamblers fallacy is the most overtly described bias in which students choose to exert control, especially in relation to SBAs and the pattern of the answers (e.g. ABCD or several of the same letter). The nature of this bias however is that recognising these patterns, and actively seeking not to be influenced by them, does not mean that the bias is not in fact in effect.

The nature of the SBA examination is such that biases were described (such as confirmation bias, anchoring bias and representativeness bias) but most commonly within a dialogue around the difficulty of avoiding these biases directly. Where students talked about directly seeking to control biases it was around recognising that they had rapidly jumped to a hypothesis and seeking to actively review and challenge this hypothesis formation. Many noted that they knew the biases existed but that within the pressured environment of the examination they would not necessarily be consciously thinking about them, even where they did take a mental step back and reconsider the initial thoughts they had.

Most data in the focus groups pertained to SBA questions. Where OSCEs were discussed the control of biases was linked to communication skills such as the use of open questions for as long as possible to prevent the risk of confirmation bias or early closure emphasising the primacy of communication skills within the reasoning process.

It can be argued that the pattern recognition element of SBA is actually a heuristic, enabling students to rapidly reach a conclusion. If we recognise pattern recognition, and its inductive nature, as linked to expert thinking we must acknowledge that this has a positive element to

it. This raises the question regarding how we consider heuristics and biases and how we articulate them in our teaching. This issue will be considered further in the implications for clinical education discussed in chapter 10.

In conclusion biases are recognised by students as common within the studied assessments, and there is some evidence of students actively seeking to manage these, however the qualitative data would suggest that the quantitative data overestimates this. This may be that students answered the question with respect to their awareness of biases, rather than specifically that they sought to manage them, within the assessment process.

9.5 Summary

In summarising this chapter it is important to revisit the original questions regarding the aim of the convergent analysis.

1. To what extent do the two datasets (quantitative and qualitative) converge or diverge?
2. Do the qualitative findings materially add to the understanding of the quantitative results?

Regarding the first of these questions this has been explored in relation to each of the ten questions with the narrative highlighting these areas whilst providing an interpretation of the results as a whole.

Regarding the second of these two questions it is clear that the qualitative results materially add to the understanding of the quantitative results and allow us to gain a better understanding of the research question.

These findings will now be considered further in the discussion chapter (10).

Chapter 10 – Discussion

10.1 Introduction and Chapter Summary

This chapter seeks to consider the results from chapter 9 and apply them to the research question and the areas previously explored within the narrative review.

This study aimed to answer the research question as to how students undertake the process of clinical reasoning in primary clinical assessments.

It is important to restate the key findings, as discussed in chapter 9 as these form the basis for the discussion and the wider implications of this work. These are summarised below:

- Students felt that their thinking was affected by being in an examination however it is likely that this is linked to general anxiety rather than a true reasoning change.
- Demographics of the patient and, to a lesser extent, the clinical setting impact on students clinical reasoning with a tendency to lead to inductive reasoning approaches focussed on pattern recognition, particularly in SBA questions.
- Early hypothesis formation is ubiquitous in the exams under study, however there is some variation in how it occurs due to the nature of the examinations themselves.
- Abduction best describes the process of reasoning being undertaken by students explaining the slight differences demonstrated between the two assessment types.
- There is some evidence for students getting to answers without certainty of the process however this appears to link to satisficing in a position of lack of knowledge rather than being linked to the cognitive continuum.
- Satisficing occurs commonly but is multifactorial in nature, particularly around knowledge and time pressure.
- Students are aware of biases, although there is confusion with heuristics. The evidence that they actively seek to manage them is less clear.

This chapter will discuss these key findings, alongside the existing literature, with clear discussion regarding the relevant contribution to new knowledge provided by these results.

The chapter will initially revisit the elements of the quantitative results that were not explored in the convergent chapter such as any notable differences between specific clinical

groups, ethnicity or gender. These need to be considered here where there may be implications for the results overall and also the areas for wider discussion around their implications for assessment, education and policy.

The next section will then review the initially constructed cognitive framework and undertake a considered revision of this in light of the findings of this study. This will be broken down, as in previous chapters, to the three core areas of external context, cognitive continuum and closure.

Finally the chapter will go on to consider the implications for assessment and education, with consideration of the implications for educators, policy and research, before concluding with a summary of key points and recommendations.

10.2 Revisiting the quantitative results

It is important to revisit and consider the quantitative results; specifically clinical role (physician associate student v medical student), ethnicity and gender. There were some differences noted within the data which need to be considered with respect to their implications for policy, education and research.

10.2.1 Differences in clinical role

The comparative data for the clinical roles demonstrated statistically significant differences in the answers to questions one, seven and eight. It is important to acknowledge that the quantitative data consisted of 400 data points from medical students and 195 from student physician associates so the groups for comparison were not equal.

In SBA questions the question one data suggested that PA students felt less strongly that their thinking was affected by being in an assessment however both of the means sit in the "often" range with a narrow inter-quartile range. This suggests that although there is a statistically significant difference both groups felt that they were often affected by being in an examination. There were comments in the PA only focus groups (2 and 7) which support that they also felt the pressure in examinations however, in line with the previous comments in chapter nine, this was less about how students think but actually about a change in performance, and cognitive ease, due to the situational pressure they felt within the examination setting.

Questions seven and eight are both linked to pattern recognition. In question seven, which specifically asks regarding pattern recognition, we see the means for both groups to be sat within the “often” range of the Likert data scale. In question eight the mean for medical students sits in the sometimes range and for student PAs in the rarely range however the means themselves sit less than a single figure apart suggesting the difference is not marked as previously noted in the quantitative results chapter.

The OSCE results see a similar pattern. Where there are statistical differences in the means for questions two, five and seven the spread of results, in terms of the interquartile range, is identical. For question two the impact of demographics in the question was slightly less for the student physician associates with the mean sitting in the “sometimes” range versus “often” for the medical students. This needs to be set against the qualitative findings which noted that this was less important information in OSCE than in SBA.

There are currently important policy debates regarding the physician associate role, with particular focus on patient safety linked to decision-making. It is therefore important that we explore any differences in thinking between these two groups (Ghadiri, 2020). The training of physician associates is designed to be in the medical model and therefore aligned to medical student training. On this basis we would expect to see similarities between the two groups, which is borne out in the results, where there are only differences in a few questions, and these are minor. It is important not to conflate the thinking process and the amount of knowledge. Physician Associate training is shorter than that for medical students (2 years v 5 years for most) and therefore the breadth of knowledge is likely to differ, however the results suggest that the cognitive reasoning processes, in assessments at least, is broadly the same. This difference in potential knowledge forms another contextual factor as per Durning et al (2012b), however despite the view of Schmidt et al (1990) that reasoning development is incremental, the reasoning approach in assessments does not appear to be unduly altered by this difference in training length. Whilst this study is focused on the reasoning process in assessments, it shows that in our education we are developing similar clinical reasoning processes for both medical students and student physician associates. This has important policy considerations regarding deployment of Physician Associates as well as strengthening the support for the medical model of education in this group (Jackson *et al.*, 2017). Further research is needed to explore whether this similarity is

seen in clinical practice as no other evidence exists regarding this difference, making this finding unique, as well as having important implications for the role and workforce planning considerations.

10.2.2 Impact of gender

There was very little gender difference in the quantitative results with only a single question showing any statistical difference (early hypothesis generation) and this was at the limits of statistical significance. This is in keeping with more general work on cognitive reasoning. Previous work has shown that men and women have almost identical general intelligence although men have a more variable distribution (Shibley Hyde and Linn, 2006). Given that the studied clinical students are selected, and required to demonstrate high academic ability, it is to be expected that there would be little difference between the two groups in their overall cognitive ability. It is interesting to note that women are found to have a greater mean ability in verbal reasoning, which is important to note in the context of interview processes, and may go some way to explain the gender differences seen in these clinical professions (Lakin, 2013). Exploration of this is out with the scope of this study.

10.2.3 Impact of Ethnicity

The data for ethnicity is almost identical between white and BAME individuals. The only statistical difference for SBA was noted in question eight where the BAME group results sat in the "sometimes" range whereas the white group sat in the "rarely" range although the difference in means was less than one Likert scale point. The BAME group also had a slightly wider inter-quartile range. There were four differences for OSCE however the most statistically significant difference was also seen in question eight. Cross-cultural differences in reasoning have been proposed previously (Nisbett *et al.*, 2001) however more recent work has suggested little difference when comparing cultures (Johnson-Laird, 2006). Importantly this latter work notes that the difference may lie in the strategies employed to reason rather than the underlying cognitive process with the implication for education policy that we must ensure a toolbox of approaches to clinical reasoning to ensure that all students can find an approach to suit them. This difference in strategies fits with the concept of abductive reasoning and strengthens the argument that this is the most important descriptor of the reasoning process as discussed later in this chapter (10.3.2).

As this study looked at final year students it is unsurprising that there is little difference as the students will have been exposed to our educational and assessment processes for 2 or 5 years (depending upon the individual student), all based on best practice in medical education. It is expected that students will therefore have learned to make decisions based upon the education we have delivered, an effect that has previously been noted in school settings (Zhang *et al.*, 2019). Support for this "training" effect can also be seen when UCAT (UK Clinical Aptitude Test) scores for admission to medical school are considered where the evidence is that BAME applicants have lower UCAT scores than their white counterparts (Brown *et al.*, 2023). On this basis the fact that at the end of the education process there is little difference between the two groups emphasises how the education and assessment provision by the school has a significant effect on the way students will reason This has important implications for education as it emphasises that the teaching of reasoning must be considered as the curriculum is developed.

10.3 Revisiting the Concept Framework

The developed concept framework from Chapter 2 is unique in the way that it was produced from combined evidence from across multiple professions. The results from this study now allow a revisiting of the concept framework and the opportunity to consider whether the research results continue to support it as representative of the clinical reasoning process, specifically within assessments, and whether the results suggest that revisions are required. The revision of the framework will be considered using the same headings as previously i.e. context (external and internal), cognitive continuum and closure. The original concept framework, as derived from the narrative review, is shown in figure 3.

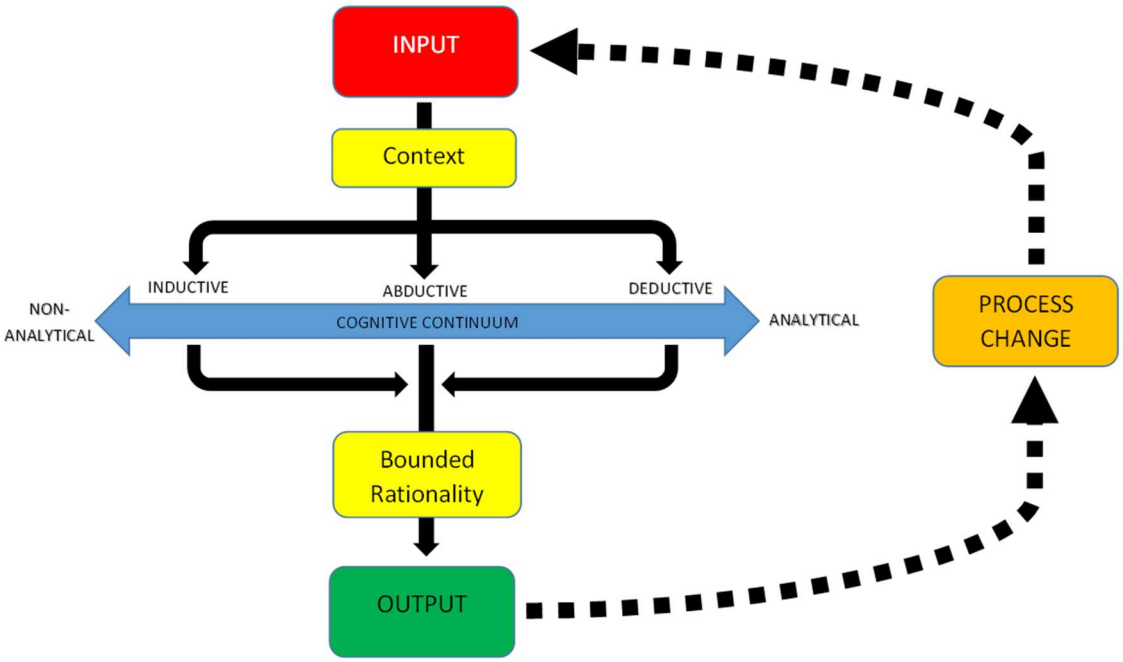


Figure 3: A conceptual model of reasoning based on the literature evidence

10.3.1 The impact of context

The impact of context has been demonstrated to be extremely important; however it is clear that it needs to be defined more clearly and explicitly. It is more appropriate to change the term "context" to that of situated cognition, which draws from situativity theory, that postulates that what we know and how we think and learn is located in the experience that we have in all its elements (Durning and Artino, 2011). Within the clinical encounter the three primary inputs to the decision-making are physician factors, patient factors and environmental factors. If we apply these same areas to assessment, we must change this slightly. For physician factors we should state candidate factors, for patient factors we should state question factors (encompassing the question content) whilst environmental factors remain the same (figure 16). Candidate and environmental factors represent external context, and question factors internal context in relation to the original groupings. This definition of situated cognition provides a better guide than the single word "context". In this study one specific environment was studied however, even so, the situated cognition was altered, particularly for the OSCE, as it was impossible to recreate totally identical experiences for these students. This is analogous to Croskerry's observations in 2009 regarding the difficulty of recreating decision contexts (Croskerry, 2009a).

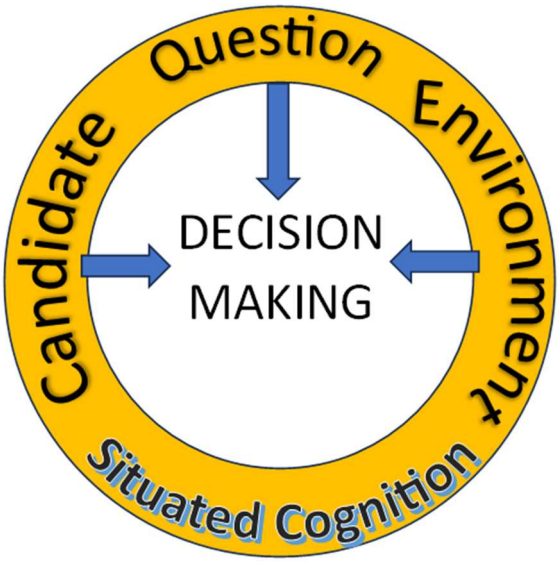


Figure 16: Situated cognition factors in assessments.

The elements of situated cognition help us to recognise how different question factors lead to different thinking through changes in elements such as the clinical setting or demographics (question factors), time pressure in the examination (candidate and environmental factors) and the way the question is designed, leading to primacy of certain decision-making processes (inductive pattern recognition in SBA). Using situated cognition rather than context provides a framework which can enable us to break down the concept into more constituent parts. This has important implications for education on clinical reasoning as it provides a structure for us to teach students about these impacts and how they affect their thinking processes, the way that different clinical environments will influence them, the risks of falling into the trap of representativeness bias and, hopefully, better allow us to prepare them for clinical reasoning in assessments and the clinical environment. There is significant value for those developing programmes of assessment through this change as the use of this model enables us to consider each of these factors and how they may be affecting student thinking. Use of this concept allows us to look at how we design assessments that vary each of these factors, with the aim of driving different reasoning processes in our students, as well as how we can create an appropriate suite of assessment processes that lead to us being able to better assess different facets of the reasoning process more effectively in a programmatic approach. This provides a direct

challenge to us as educators to ensure that assessments are considered in relation to this concept to benefit students' cognitive development.

The data has shown that situated cognition will impact the reasoning process as well as the decision closure with satisficing being directly influenced by these factors. The positioning of this concept above the cognitive continuum in the original concept framework does not therefore hold true. This fits with the previous work of Durning et al (2013) and Gibson (2000) as the candidate factors such as attitude and the effect on reasoning of the environmental and question factors, impact on the entire reasoning process. It is more accurate to state that situated cognition encompasses the entire clinical reasoning process and influences it throughout. In revisiting the concept framework we therefore have to place this to encompass the process in its entirety. This better reflects the true impact of all elements of situated cognition in the process and enables us to emphasise its impact on the cognitive process.

10.3.2 Considering the cognitive continuum

The combined results from the cognitive continuum show a strong preference for pattern recognition but this tendency is more pronounced in the SBA compared to the OSCE. This fits with previous work demonstrating that medical students reason in the same way as doctors (Neufeld *et al.*, 1981; Lopes *et al.*, 2018). In both exam types the data clearly indicates that whilst students will tend toward using inductive decision-making approaches this is data driven through patterns and keywords and, when these are not in clear evidence, they will be more deductive in nature. If educators are to provide higher quality education to students on clinical reasoning it is therefore more appropriate that we no longer talk about purely inductive and deductive decision-making. We must instead recognise clinical reasoning as abductive (which will vary based on the information provided) and utilise this as our starting point for any assessment or clinical encounter. Whilst both dual-process theory (Kahneman, 2011) and the cyclical model (Marcum, 2012) have a place in helping to guide students early understanding of clinical reasoning we need to ensure that by the end of training they understand the inherent fluidity in the process and its variety of potential approaches captured within the concept of abductive reasoning. Compared to inductive or deductive decision-making approaches abduction differs in that it recognises that in order to either induce, or deduce, our answer we must make an initial

leap to either conclude that our observations fit a specific rule or, that based on a series of individual observations, that we can induce a likely conclusion. In this case we might state that “This patient has had rectal bleeding, this is one symptom of possible bowel cancer, it is possible that this might be bowel cancer but there are other possibilities”. The key tenet of abduction is that it accepts that there cannot be certainty as other possibilities can exist, arguably returning to the post-positivist paradigm as discussed in chapter three and supporting the use of mixed methods to explore clinical reasoning (Flick, 2018). This provides an explicit challenge to educators regarding the teaching of uncertainty. Abduction, by its very nature, introduces uncertainty to the clinical reasoning process. Uncertainty is a natural concept within medicine (Wellbery, 2010) yet the natural approach to education is that of an expectation of diagnostic certainty. As educators we must embrace the concept of uncertainty and ensure that students are taught how to manage it as it is inherent in clinical reasoning. Uncertainty must form a core part of clinical reasoning curricula if we are to produce clinicians with an appropriate skill set for the working environment. Abduction explains the initial generation of hypotheses, which this study has demonstrated, and it allows us to formulate a possible cause from which we will then use a variety of reasoning processes to develop our answer. These will include induction and deduction but can also include probabilistic and instrumental reasoning depending upon the problem to be solved. This concept allows us to explain the differing nature of the reasoning process between different assessment types and unify them effectively. By utilising the abductive model we can move students away from a binary approach to a more reflective one and allow them to engage with their own preferred thinking style within a single theoretical framework. Similarly it allows us to bring together the diagnostic and therapeutic decision-making literature and emphasise the narrative basis of decision-making linked to varying information sources (Hunter, 1996; Greenhalgh, 1999).

When considering a revised concept framework based on the findings from this study it is arguable that there is no need to separate early hypothesis formation out as it forms a fundamental part of the abductive method. It is important though that the concept framework is also an educational aid. Separating it out will allow us to emphasise its part in the process and permit students to engage with the early hypotheses they form. Through

this inclusion, out with the simple statement of abductive reasoning, we can add value in the revised concept framework rendering it unique in explicitly stating this concept.

10.3.3 Closure

The results clearly demonstrate that bounded rationality and satisficing occur in the clinical reasoning process in assessments, justifying the statement in the narrative review that they must be included in any such concept framework. In this respect the original concept framework remains consistent with the findings. There is a question as to whether we should use the term bounded rationality or satisficing to best reflect the process that is occurring, accepting that the difference between the two concepts is subtle.

Satisficing accurately describes what is occurring at the later point of the reasoning process in assessments. Satisficing is defined as a decision-making strategy that aims for a satisfactory or adequate result, rather than the optimal solution (Simon, 1955). This fits well with the qualitative results, particularly within OSCE examinations and the mark-scheme driven approach. Bounded rationality is a more over-arching term that describes how our rationality is limited during decision making meaning that rational individuals will seek to satisfice. If we consider the answers to the first question regarding the impact of being in an assessment, and the fact that it affects the reasoning process, the results suggest that bounded rationality applies to the entire process rather than just the closure. If students are overly stressed, then this will affect the individual's ability to act rationally and might lead us to the issues articulated in the focus groups such as students "second-guessing" themselves or cognitive inertia. In the SBA papers, as students started to talk about time pressure, this represents one of the core elements of bounded rationality, namely that decisions have to be made within a set time scale. In this respect satisficing is more specific to the closure of the reasoning process whilst bounded rationality, and the ability of the individual to satisfice effectively, becomes a factor affected by the environmental aspect of situated cognition. This approach of considering bounded rationality to be a more encompassing concept fits with the statement of Young (2014) that when cognitive load exceeds working memory capacity performance is impaired. By taking this approach it also assists us in understanding the way students considered their answers to the first research question around how being in an assessment affected their thinking. As suggested in chapter nine this was not truly about the process of clinical reasoning but is clearly an aspect of bounded rationality and

may lead to students ascribing differing values to the various aspects of the exam question with a potential differing outcome. The “second guessing” that is described is clearly an indicator of this difficulty and reflective of the students struggling to decide which variables are the most important in determining the answer. In this respect representing bounded rationality as an overarching concept within which the entire decision sits is the only logical approach whilst satisficing is a very specific element of decision closure. On this basis both need to be represented within the revised concept framework but in different ways. This also re-emphasises the need for educators to ensure that the concept of “controlled uncertainty” is taught to students as part of their course. The General Medical Council talks about uncertainty of diagnosis in outcomes for graduates (GMC, 2020) and in its generic outcomes (GMC, 2017) however the focus is on its management and communication rather than managing the individual practitioner’s own uncertainty in the process. Policy needs to be changed to explicitly recognise the individuals clinical reasoning uncertainty and the need for undergraduate and postgraduate educators to support the development of this core skill.

10.3.4 A revised conceptual framework

When we consider the above conclusions from the work it leads us to revise the previous concept framework. The revised framework is designed to emphasise the all-encompassing nature of both bounded rationality (as this defines the frame of any decision and the likely value placed on the data needed for the decision) and situated cognition (as this provides a more nuanced approach to considering those elements of the process previously labelled as context. The revision also emphasises that the process of reasoning is abductive in nature which will then encompass any, and combinations of, a number of different decision-making processes which will vary depending upon these key external conditions. The revised concept framework for reasoning during assessment is shown in figure 17.

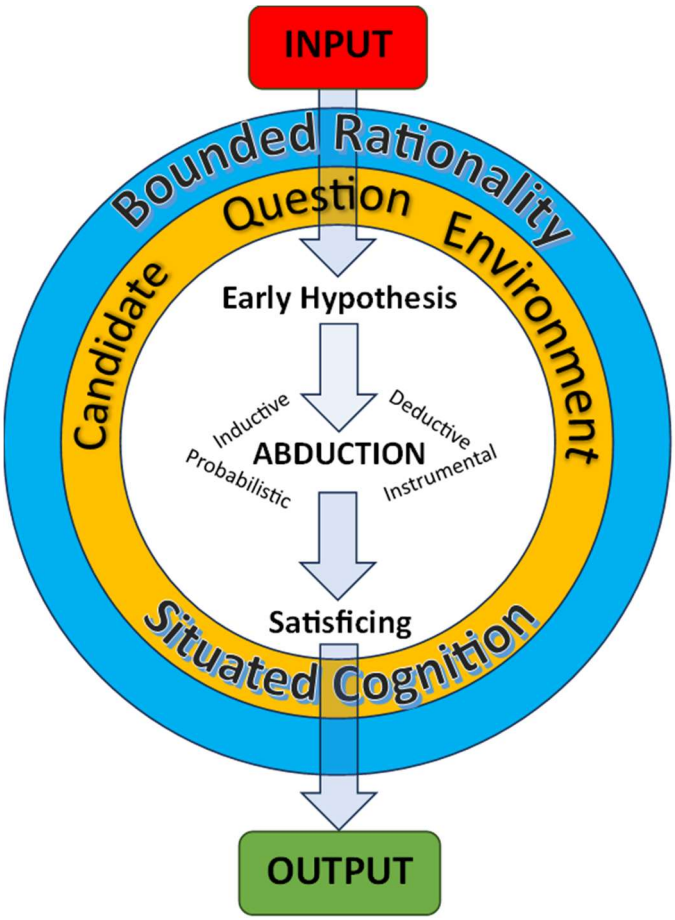


Figure 17: The revised concept framework.

The revision presents a framework to support educators in delivery of teaching to students regarding the approach to clinical reasoning. The use of abduction allows us to describe the process more like a toolbox of methods to be used as the data allows, a much more realistic approach than a binary model. With bounded rationality acting as more of a frame to the decision, satisficing becomes the more appropriate term at the closure of the process leading to the ultimate decision output.

As educators we will have to address the challenge that abduction makes to evidence based medicine within education, as noted by Upshur (1997) but also emphasise that clinical decisions are always about the best evidence gained from large studies applied to a single unique patient (Ahlsen *et al.*, 2018).

If we wished to generalise this in decisions made in environments away from assessment the only required change would be changing the elements of the situated cognition to clinician factors, patient factors and environment. This provides us with a clear and simple framework against which we can consider the various elements of our teaching and

assessment with respect to reasoning and educators can use this to plan, deliver and provide scaffolding to, appropriate education. More research will need to be done to confirm validity of the framework in other non-assessment situations however the approaches used in this research can act as a guide to do this.

10.4 Implications for Assessment

The results provide a number of clear implications for the way we undertake assessments in clinical courses.

10.4.1 Time pressure

This study suggests that time limits in examinations may lead to a change in cognitive ease with a potentially adverse effects on effective clinical reasoning, or at least driving students toward induction, rather than more deductive, reasoning approaches. Time pressure in assessments directly impacts bounded rationality and leads to a change in the way that students approach the reasoning process. It needs to be considered as to whether applying time limits on assessments, especially written ones, is actually leading to cognitive error and consequently being unfair to those sitting it. Even if we do not consider it to be unfair this study suggests that it is having an effect on the cognitive process, analogous to previous findings that cognitive processes are affected by both parties in the consultation (Durning *et al.*, 2013). In this case the two parties are the candidate, and their anxieties, and the time pressure from the question paper constraints. If we recognise the truism that "assessment drives learning" it can be implied that the pressure that students are feeling in the examination is leading to them choosing pattern recognition, induction, as their preferred approach to clinical reasoning, driving them away from other approaches as per the abductive method. Assuming an assessment is closed book, and thus the performance is still focussed on the knowledge of the individual, there is a clear argument to remove time limits, or at least make the time available of a length that ensures this situational factor is removed, or at least altered. This can reduce the impact that is has in driving inductive clinical reasoning. Given the evidence that changing elements of context can alter clinical reasoning (Durning *et al.*, 2012b) it may be that changing things in this way can allow us to facilitate abduction, as determined to be appropriate in the concept framework. Studies have demonstrated that removal of time limits in examinations is beneficial to students with

specific learning difficulties (SPLDs) but has little effect on those without (Runyan, 1991) although it should be noted that the study looked purely at the results not the cognitive process. Whilst removing all time limits is unlikely to be viable assessors need to consider the use of significantly extended time periods, certainly for written exams, to provide the opportunity for students to choose other approaches from the "abductive toolbox".

Time considerations within OSCEs are potentially more problematic due to them being rotational multi-station assessments. There is a clear mandate for those designing assessments to ensure that the task set is manageable within the time provided. This study demonstrated students still taking an inductive approach to OSCE stations due to the time pressures and focussing on "building points" rather than being so concerned about any diagnostic or therapeutic decision that may be required. Given the previous findings that this is in conflict with the concept of effective diagnosis and expertise (Schuwirth, 2009), assessors need to consider whether the current structure could be altered through the station construction, time, or weighting of aspects of the assessment to try and change this. This would have the potential to bring the assessment more into line with the reality of a normal clinical encounter. Careful management and auditing of OSCE stations will be crucial to ensure that they are fair and manageable and, with universities in the UK now delivering the clinical and professional skills assessment (CPSA) of the medical licensing examination for the GMC, it should be considered as to whether such post-hoc evaluation should become a policy expectation.

10.4.2 Question order

Issues of exam question order in the single best answer papers present another implication for our assessments, particularly in very high stakes assessments that clinical students will have to undertake (Medical Licensing Exam, Physician Associate National Examination). Previous authors have experimented with randomly ordered questions for different students in the same examination and found no evidence of detriment to performance, with the suggested benefit of improved examination integrity, however it is unclear as to the implications of poor performance in the test under study (McLeod *et al.*, 2003). This study showed that commencing an examination with a series of difficult questions could unsettle students, whereas starting with easier questions could create an earlier feeling of cognitive ease with clear implications for the reasoning process. Given that most SBA assessments are

now taken digitally assessment leads should consider whether returning to non-randomised assessments will ensure that all students are subject to the same cognitive challenges and thus ensuring equity in the assessment process. It is important to recognise that situated cognition has an impact here as difficulty is individually subjective.

This issue is unlikely to be easily solvable with OSCE. The idea of all students doing the same circuit in the same order is unrealistic given the cost and assessor intensity of these examinations.

In both SBA and OSCE our standard setting approaches focus on the borderline candidate, with the commonest approach for standard setting SBA being Angoff, and for larger OSCE groups borderline regression (Ben-David, 2000). It is likely to be the borderline candidate who is most affected by these issues and so assessment leads need to consider them seriously, with appropriate policy guidance as appropriate, so as not to disadvantage this group.

10.4.3 Abduction and its implications for assessment

The results of this study demonstrate that students favour an inductive pattern recognition approach to clinical reasoning in the studied assessments, particularly SBA. Recognising that the clinical reasoning process is abductive (Ward *et al.*, 2016), as clearly articulated in the revised concept framework, may reassure us somewhat that the choice of induction is simply a strategy within the abductive process however we need to ensure that our assessments provide opportunities to explore the other reasoning approaches. We can liken this to the point made in a 2005 article exploring the methods of fictional detectives which concludes that an ideal clinician will utilise all of the differing approaches seen in these sleuths (Rapezzi *et al.*, 2005).

The difficulty we have is that this study shows that the studied assessments drive inductive reasoning. This fits with the findings of previous authors in Emergency Departments and walk in centres (Gruppen *et al.*, 1988; Pelaccia *et al.*, 2014). The recommendation that only relevant information is included in question stems, or OSCE cases leads, to some of the issues seen within this study. Ethnicity is the obvious example as previously discussed. There is an argument that ethnicity should be included in every question as it is something that you cannot “unsee” and doing so would be an attempt to try and remove the racial

stereotyping that occurs when it is only included as it directly impacts on the decision to be made. In reality this will simply add to the risk of "white noise" within the questions and, in all likelihood, fail to achieve what we wish it to. This is discussed further in section 10.5.4.

If we accept that assessing clinical reasoning is about us testing that students can apply different reasoning approaches, we need to ensure an appropriate plurality of opportunities to do so providing a clear challenge to educators and current assessment schemes. This will be discussed more below.

10.4.4 Workplace based assessment

This study focussed on the SBA and the OSCE, but the findings emphasise the vital role of workplace-based assessment in helping to ensure a rounded assessment picture. The results of the question linked to "gut feeling", associated with the satisficing concept within the concept framework, and the fact that this was seen to be less significant in OSCE is contrary to previous authors findings in General Practice (Stolper *et al.*, 2009). This emphasises the need for educators to provide more assessments in "real world" settings but directly considering the clinical reasoning approaches of the individual. The need for real patients to provide stimulation for these elements of the reasoning process, and potentially introduce the impact of biases more into the assessment, is vital if our schemes of assessment are to provide a rounded picture of a student's true clinical reasoning ability. Given the impact of situated cognition we may need to revisit these assessments to ensure that they can effectively consider the impact of the environment on the decision and a significant challenge will be to ensure that our clinical educators are effectively trained to recognise and discuss these issues. In doing so our suite of assessments can be made more fit for purpose.

10.4.5 Creating assessment schemes for clinical reasoning

The discussion above highlights the need for our assessments to be designed to ensure trainee clinicians are presented with opportunities to utilise multiple reasoning approaches and, as educators, we must openly acknowledge the limitations of SBA and OSCE as our primary approaches. In acknowledging this it returns us to the concept of programmatic assessment and the need to consider the differing assessments through an entire programme and, as this study shows, consider the reasoning processes being utilised. This is

crucial if we are to produce clinicians with an effective "toolbox" of reasoning strategies to aid them in the workplace. The revised concept framework developed in this work can be used to support educators in developing schemes of assessment by considering how each framework element impacts on any given assessment approach and thus the different elements of, and approaches to, clinical reasoning that are being evaluated.

Using the concept framework can allow educators to consider alternative assessment approaches within a suite of testing across a clinical course. This is not about removing SBA and OSCE from the approaches that we use but recognising that they must be part of a scheme of assessment in which clinical reasoning development is also considered. Programmatic assessment views the use of multiple formative assessment review points to be crucial and it may be that we can be more innovative in some of these to allow differing aspects of clinical reasoning to be explored (Van der Vleuten *et al.*, 2012).

The use of very short answer questions (VSA) has been suggested as an alternative to the traditional SBA with studies demonstrating both their validity and benefits in efficiency over more commonly used short answer questions (SAQs) as they can be machine marked (Sam *et al.*, 2016). Due to their construction these have the same flaws as the SBA in so far as they have a similar stem construction and therefore are liable to drive induction. The removal of answer provision may drive some alternative thinking processes as they can no-longer be used as a proxy for early hypothesis formation however they are not different enough to provide the whole answer.

There is evidence that students themselves prefer structured oral examinations to test clinical reasoning, believing them to be fairer than workplace based assessments and OSCE (Tolsma *et al.*, 2024) and some work shows that key competencies, including clinical reasoning can be assessed through this approach (Pernar *et al.*, 2020). There is no doubt that an oral examination allows more nuanced ability to assess clinical reasoning by considering all of the core aspects of the process as set out in the concept framework however caution must be taken to ensure that these remain appropriately structured and thus subjectivity reduced (Shenwai and Patil, 2013). Inclusion of such assessments may allow us to evaluate student thinking more effectively as we can seek descriptions of the process they are following and these may be more effective than the "double OSCE" stations described by Durning (Durning *et al.*, 2012a). In these approaches an initial station

requires students to undertake a decision-making process followed by a second station immediately after in which the candidate talks through their decision-making approach. It is notable that a scoping review of assessment methods, and their value in measuring the different decision-making components, would suggest that the combination of the two gives much greater evaluation of decision-making across all components (Daniel *et al.*, 2019). The approach of oral discussion versus the "double OSCE" approach needs objective comparison regarding their assessment of clinical reasoning. It may be that aligning the oral case study and double OSCE approach can allow useful cross over between OSCE, a simulated case study, and evaluation as part of workplace-based assessment. Similar work has been done evaluating students written description of their decision-making processes in the OSCE environment (Siegelman *et al.*, 2024), although in the work by Daniel et al, written approaches score lower for their evaluation of the different clinical reasoning components. In either case these are talk aloud approaches and therefore subject to the same risks around the ability to assess a process that, if it reaches an appropriate conclusion, cannot be said to fit easily into a right or wrong binary decision.

Exploring formative assessment opportunities developed around the primary summative assessments may give us an opportunity to innovate and trial different approaches. The concept framework developed in this work can be used to support educators in developing their schemes of assessment, whilst researchers in medical education need to evaluate these approaches in order to support the development of appropriate guidance.

Finally it has been noted that workplace based assessment needs to be considered carefully as part of our overall assessment strategy to ensure the educational value to the student (James *et al.*, 2009). We must be certain that whichever workplace-based assessment tool, or tools, are utilised that there is consideration of the clinical reasoning element of the interaction within it. It may be that we need to have two separate approaches to workplace-based assessment with one based upon the doing (history construct, examination etc) whilst the other focusses on the cognitive process. The concept framework developed in this work can facilitate the construction of such a tool which will then need appropriate evaluation in the clinical environment. These will clearly require significant staff training to be effectively delivered, however assessment within the "real world" is likely to deliver the most realistic simulation of reasoning knowledge and application.

10.5 Implications for Clinical Education

Providing appropriate education on clinical reasoning is crucial if we are to try and improve patient safety. With a systematic review showing that 36.5% to 77% of diagnostic errors were linked to cognitive biases there is a clear imperative to improve the teaching of this topic in our curricula (Saposnik *et al.*, 2016). Previous work has highlighted the need for clinical reasoning to be embedded in curricula and assessed effectively (Kononowicz *et al.*, 2020) and this study emphasises the limited approaches currently taken by students directly challenging the current status quo.

10.5.1 Self-regulated learning and Metacognition

Much of the considerations for clinical education can be framed around the concepts of metacognition and self-regulated learning. Self-regulated learning (SRL) is a process by which students control their own learning through cognitive and metacognitive strategies and is often considered as a cyclical approach of three phases, forethought, performance and self-reflection (Artino and Jones, 2013). In this cyclical form it can be considered analogous to the concept framework developed in this study with forethought linked to hypothesis generation and situated cognition, performance to the abductive reasoning process, and self-reflection to bounded rationality and satisficing. This is to be expected considering that both the concept framework and SRL are focussed on an individual's approach to task completion. Most of the literature around SRL focusses on the diagnostic decision-making process, and the assessment of it (Goldowsky and Rencic, 2023), with little focus on assessment of clinical reasoning as set out by the operational definition in this work, however microanalytics presents an opportunity in this area. Microanalytics has potential, within the concept previously discussed in section 10.4.5, of combining talk aloud style processes with existing assessment approaches. Within formative and summative assessments, particularly in the workplace, the development of structured, process focussed, interview questions can facilitate clinical reasoning assessment and metacognitive evaluation both in teaching and assessment (Cleary *et al.*, 2016).

Clinical educators have to consider the entire clinical reasoning process in patient care, as set out in the operational definition of clinical reasoning within chapter 2 of this work, if we are to unlock the potential of SRL and improve student metacognition as a whole. The focus

on diagnostic decision-making is understandable however it should equally be applied to the therapeutic decision-making process. Further research into the development of microanalytic assessment and its evaluation is needed to enhance the opportunity to integrate this approach into programmatic assessment schemes. Educators need to redevelop formative and workplace-based assessments to enable these approaches to be developed and facilitate greater breadth of metacognitive evaluation.

10.5.2 Diagnostic stewardship

Diagnostic stewardship is a term that was coined by Daniel Morgan in relation to the appropriate ordering, performing and interpretation of diagnostic tests, in particular linked to antibiotic stewardship, aiming to reduce unnecessary usage (Morgan *et al.*, 2017). More generally diagnostic stewardship has been described as aiming to ensure the right test for the right patient, in order to achieve the right management (Fabre *et al.*, 2023). Despite its obvious common sense diagnostic stewardship remains a poorly understood term. Logically diagnostic stewardship has to begin with effective clinical reasoning of patient diagnosis as, in order to ensure the first step of performing the most appropriate test, we must ensure that this initial step can be taken effectively. If our current teaching and assessment processes do not focus on developing clinical reasoning beyond the diagnosis, then we are failing to adhere to this concept and provide students with a comprehensive toolkit of reasoning approaches with which to apply their knowledge effectively. This was noted in a 2018 policy briefing (Graber *et al.*, 2018) . This is important in the way that we teach clinical medicine, particularly in the clinical environment, and especially given the primacy of pattern recognition, and situated cognition, impacts upon the decision-making process. There is good evidence that, in the active clinical environment, we fail to manage this effectively with a number of studies demonstrating the unnecessary use of investigations in different circumstances and the potential for cost savings and improved clinical care (Hogg *et al.*, 2005; Amin and Wang, 2009; Vrijsen *et al.*, 2020). As educators we have a responsibility to put clinical reasoning at the centre of what we do as the evidence shows that it is good for patients, and health services, and this has to start with how we design our teaching and assessments to meet this need. Specifically we need to recognise diagnostic stewardship is not simply about making diagnoses but the continuum of decisions to be made as per the operational definition of clinical reasoning in this work. This creates a direct

challenge to educators to develop their own understanding, and teaching, of this concept in order to improve the clinical journey and outcomes for patients. Clinical reasoning directly impacts resource management and thus policy makers also need to elevate the requirement for this to form a core part of medical training at all levels. The concept framework can be utilised to frame this process and allow educators to support students understanding of each element and improve their metacognition.

10.5.3 Situated cognition

Educators need to ensure that situated cognition is a fundamental part of our teaching practice. The decision-making process cannot be seen as isolated from the situation in which it occurs, and it is important that we stress the impact in assessments of the three key factors as defined in this work. Bringing situated cognition, as part of the clinical reasoning process, into workplace based assessments is an important aspect of developing skills in this area (Rencic *et al.*, 2020). We must also ensure that our clinical education considers, and openly discusses, the situated cognition elements that influence any particular decisions. With the students in this study (medical and physician associate students) rotating through multiple different clinical areas this is crucial to support diagnostic, and inferentially, therapeutic, stewardship. The results demonstrated that the clinical setting of a question seldom had a key role in the approach to a diagnostic decision, but it did for management options. Demonstrating that this is true for assessments, we can infer that situated cognition will impact the reasoning process for diagnosis in the clinical setting where a patient may be at risk of harm if decisions are not made appropriately both in terms of diagnosis and signposting. Ultimately the goal of developing clinical reasoning within curricula is to reduce these errors in practice (Saber Tehrani *et al.*, 2013). Integrating this can improve the development of effective diagnostic skills and holistic patient management (Graber, 2020). The situated cognition model can aid us in building fit for purposes assessments but needs to be considered in education to support our diagnostic stewardship. Educators need to ensure that this is a standard part of the assessment process to ensure student clinical reasoning development that is fit for purpose. More research using microanalytic techniques can allow us to further understand the impact of situated cognition and support students in their development.

10.5.4 Teaching approach and bias

The results, as they pertain to the primacy of pattern recognition within the thinking process in assessments, need to be considered by educators, especially in light of the comments within the focus groups that it is the way that students are taught that drives the primacy of this thinking process. Just as it is often stated that assessment drives learning, it can be inferred from the data that teaching drives the preferred clinical reasoning approach and that at present it tends toward driving inductive clinical reasoning by the teaching of conditions and the formation of mental illness scripts. Educators must consider the teaching that is being undertaken and how it can be developed to provide a plurality of clinical reasoning approaches to be learned.

This reimagining of our educational approach need not all be tutor led. Work with student pharmacists has demonstrated that students can provide feedback on clinical reasoning that is just as effective as that provided by tutors (Shabanowitz *et al.*, 2024). If educators build teaching programmes around the process of reasoning and the impact of the elements of situated cognition, as well as training students how to recognise and feedback on it, we can create peer led processes to drive clinical reasoning development which is beneficial for both students and educators. We need to involve the students themselves in providing evaluation of clinical reasoning in different ways if we wish to develop their understanding of different approaches to managing it.

The use of “typical pictures” in education is a challenge to effective learning and development of clinical reasoning skills. Previous work has challenged the issue of problem based learning having a biomedical focus, rather than a more patient focussed one, despite the latter being crucial for effective clinical reasoning (MacLeod, 2011). The “typical picture” approach will drive students toward an inductive reasoning approach as well as increase the risk of representativeness bias which can lead to students taking inappropriate assumptions from the exam paper to the clinical environment.

Pattern recognition, along with the guidance that we ensure that we do not include more information than is required in questions, risks our attempts to diversify the clinical curriculum by driving representativeness bias. In a diverse nation such as the UK it is especially important that we seek to ensure that our teaching and assessment does not

reinforce stereotypes, yet there is a clear risk that we do so in our teaching (Gishen and Lokugamage, 2019). As this study shows, when ethnicity is seen in a question students immediately look for a diagnosis that is "linked" to that ethnicity. In reality, the fact that something is more representative does not necessarily make it more likely (Kahneman and Tversky, 1973) however our teaching often continues to focus on these "typical" patterns and reinforces this bias. Educators must grasp the challenge to manage issues of bias, however bias is recognised as a "wicked " problem (a complex issue with an unknown number of potential solutions) (Horst and Webber, 1973) and single interventions are unlikely to be successful (Gopal *et al.*, 2021). In addition there is very little evidence of effective de-biasing strategies, even where biases are identified, presenting difficulties in transacting this. The language used by students in the focus groups in this study showed that the term bias is used interchangeably for both negative biases and heuristics, the latter having been shown to be a core part of the reasoning process and linked to the satisficing process, particularly in experts (Kahneman *et al.*, 1982). Educators need to openly acknowledge heuristics and biases and link them into our clinical teaching so that they are "out in the open", demonstrating them in clinical scenarios to facilitate understanding. In doing so we can acknowledge them in our teaching and guide students' bias awareness and heuristic development, with a view to improving their understanding and increasing the scope of the cognitive toolbox we provide them with (Feufel and Flach, 2019).

Educators must introduce education regarding cognitive biases early in clinical education if we are to try and reduce individual susceptibility to these. A Dutch study looked at resident physicians susceptibility to anchoring bias and discovered that where salient discriminating features were present, i.e. patterns to recognise, anchoring bias was more prevalent, however it was less prevalent in those with greater knowledge of these features (Mamede *et al.*, 2024). Recognising that we have students at the very early stages of their clinical journey it is incumbent upon us that instead of simply focussing on single conditions we take more of a comparative approach based on symptom presentation and distinguishing features to help reduce anchoring.

Educators must work to develop student understanding of the difference between heuristics and biases. This study showed the ubiquity of the word bias when students were actually describing their mental shortcuts as linked to pattern recognition (I am biased that a

patient with presentation X will have condition Y). Separating the two gives us the ability to support positive heuristics whilst emphasising the risks inherent in the more negative cognitive biases. With the lack of research on heuristics further research into this area has the potential to provide a framework to support this work.

10.5.5 Satisficing

The concept of satisficing, particularly within situated cognition, needs to be further explored by educators. The feeling of a need for perfection was articulated by a number of students in this study yet the reality of clinical care is that many decisions do not lead to a definitive answer instead directing us toward differential diaganoses and treatment options which are then explored over time and through investigation. Given the known challenge of clinician burnout (Dzau *et al.*, 2018) and the link between perfectionism and depression (Smith *et al.*, 2018) educators must ensure that students are able to understand, and accept, that the cognitive process does not necessarily need to lead to a perfect answer, and that good enough is acceptable, with our aim in any clinical decision being to provide safe and effective clinical care for the patient. By bringing this concept into the open we can better prepare students for clinical practice as well as hopefully facilitating their mental health both as students and as practising clinicians (Enns *et al.*, 2001).

As educators we must articulate the difference between bounded rationality and satisficing to students in order to facilitate their understanding of the difference between the two and enable their understanding of the impact of variation in elements of the bounded rationality concept. An obvious example of this would be the way that time pressure decisions in Emergency Medicine may differ from those in an outpatient or primary care setting and how this can affect effective decision-making and the acceptable outcome. This will require curricular review and integration of these concepts into our programmes of education. Research will be needed to understand how students recognise and process these principles.

10.5.6 Technology and the future

Technology has significant potential to improve our approaches to the teaching of clinical reasoning. Generative AI has already been mentioned in this chapter and is only one of a number of technological developments with the potential to facilitate clinical reasoning

development. In doing so we must ensure that these are led by our learning aims and objectives rather than risking these approaches being “shoe-horned” into an already busy curriculum.

Generative AI has already been used to develop virtual patients who can interact with trainees to educate them effectively (Sardesai *et al.*, 2024) and can be utilised in the evaluation of clinical reasoning skills (Çetinkaya *et al.*, 2024). Virtual patients are acceptable to students and have been used in both individual and group scenarios to support reasoning development and, with a study noting that students felt they could allow them to see their cognitive errors, have the potential to be used as formative assessment tools as part of a programmatic approach (Gonullu *et al.*, 2024). Combining this with approaches such as unfolding case studies, but with the AI enabling differing decisions to lead to differing consequences, is clearly advantageous with reduced cost and lack of reliance on actor availability. Unfolding case studies have been shown to be beneficial in permitting students to practice and develop reasoning skills in a low-risk environment (Williams and Nottingham, 2022) so this has promising potential. Additional potential exists with the developing AI ability to gather contemporary data regarding students thinking patterns and provide feedback on the process. A 2021 study using a virtual patient environment allowed information to be collected on students’ information gathering approach, in particular the elements of diagnostic stewardship discussed previously (Zheng *et al.*, 2022). Feedback from such technology can provide effective feedback to learners allowing them to develop their self-regulated learning more effectively. Despite this more work needs to be done to develop the technology to evaluate more complex elements of the process such as initial hypothesis generation and diagnostic justification (Jay *et al.*, 2024), all important elements of the concept framework developed through this work.

It is worth noting that multiple uses of innovative technologies are described in literature although without clear adoption into the mainstream. Serious games have been explored and shown the potential to facilitate clinical reasoning development however they are not without the same challenges that exist for AI around which components of the reasoning process are effectively assessed (Koelewijn *et al.*, 2024). Approaches using other mainstream technology platforms such as Tik Tok may have potential but as yet remain limited in their evaluation.

Technology is likely to be assistive in clinical reasoning development going forward with the potential for us to support student development of several differing process elements. These need to be aligned to our pedagogical aims and therefore further emphasise the need to have a clear strategy for clinical reasoning development throughout clinical courses. It is notable that a 2024 paper discussed a protocol for a scoping review to evaluate gamification's impact in clinical reasoning education and it will require this level of distillation of the evidence to permit us to better understand how we can move these multiple technological approaches from single site approaches to the mainstream (Lee *et al.*, 2024). Where technology is to be incorporated directly into assessment of clinical reasoning the concept framework in this study can be used to help evaluate whether such approaches provide evaluation of all aspects of clinical reasoning or simply some elements that can be complimented by the programmatic use of other assessments approaches.

10.6 Summary

This discussion chapter has brought together the data from the two arms of the study as recommended in the convergent design. These results have then been used to revisit and redevelop the conceptual framework to ensure it is fit for purpose.

The results provide us with a clear picture of the way students think in the SBA and OSCE assessments which enables us to consider the implications of these results to both assessments and clinical education more generally. There have been a number of specific challenges to educators and researchers made in this chapter. Policy change is potentially needed if we are to realise the learning form this research. The conclusion will summarise these elements and recommend the next steps from this research work.

Chapter 11 – Strengths, Limitations and Mitigations

11.1 Introduction and Chapter Summary

As with all research work this project has both strengths and limitations. This chapter will discuss these along with the mitigations put in place to try to address the limitations.

The chapter will be broken down in relation to the differing stages of the project. Under each section it will discuss the strengths of the project at each stage followed by consideration of the limitations. It will initially consider the literature review, including the production of the concept framework, before considering the nominal group approach and then the project itself, including its overall mixed methods approach, and then the quantitative and qualitative components and outputs.

Where limitations exist, and where there has been an attempt to mitigate these limitations, this will be discussed. Finally insider research and its application in terms of both strengths and weaknesses will be discussed as a separate section toward the end of the chapter before a brief summary.

11.2 The Literature review and Concept Framework

11.2.1 Strengths

As previously noted in the literature review chapter undertaking a literature review on a topic such as clinical reasoning is challenging due to the nature of the literature concerned with the topic not lending itself well to the "gold-standard" of systematic review. This narrative review was strengthened, compared to many, by considering literature from across the clinical professions, without focussing on any single one, and permitting the production of a new concept framework for clinical reasoning. The framework is unique to this work and, through the results of the study, has been revised. The framework, along with the produced operational definition for clinical reasoning developed in chapter 1, stood up to testing and can provide a useful framework for teaching and assessing clinical reasoning within clinical education of all types.

11.2.2 Limitations

The search terms used in the review were designed to provide a suitable starting point in the literature which, combined with citation review, could enable an appropriate breadth of literature to be accessed. There is a limitation that the literature is not, by definition, exhaustive and some papers that might be considered important may have been excluded. This is mitigated by the approach to the review, including citation review, and the fact that the concept framework held up within the research process. The choice of a Narrative review was a pragmatic one, in line with the nature of the literature, and underpinned by an appropriate review methodology (SANRA) to ensure that it was able to provide a suitable background to the work (Baethge *et al.*, 2019). External review of work at completion against these criteria was designed to ensure that it met those recommended requirements and avoided potential researcher bias in the process.

11.3 The Nominal Group Approach

11.3.1 Strengths

The nominal group approach to developing the questions used in the study provides a strength to the questions used and adds validity to them. The approach ensured that the questions represented the views of a more diverse group, utilising combined expertise, rather than the question set being driven purely by my own knowledge with its inherent risk of bias. These experts also used the constructed concept framework to help support the decision-making and therefore considered it to be appropriate to the research question and the general subject matter.

The use of an online approach ensured the group could be geographically diverse and reducing bias that might occur if only those locally were involved and used a novel approach of data capture through the use of online spreadsheets that allowed multiple user engagement simultaneously.

11.3.2 Limitations

Recruitment to the nominal group was undertaken through a call for volunteers from the CReME (Clinical Reasoning in Medical Education) group from the UK. Only small numbers of individuals expressed an interest through this route so there was consequently no choice of specific individuals to make up the group. Two local colleagues with a particular interest in clinical reasoning were also invited to contribute. The group itself was therefore made up of

all those who offered to support the project which, whilst within quoted limits for such an approach, was slightly fewer than would be optimum.

The nature of the recruitment, as set out above, led to a geographically diverse group with three from university 1 and three from elsewhere in the UK, however it did not give any ethnic or gender diversity with all six being white men of UK origin. It is difficult to speculate whether this is likely to have impacted upon the outcome of the process, given the literature regarding gender differences in clinical reasoning is extremely limited, however it should be recognised as having the potential to impact on the study. The literature on ethnic variation in reasoning tends to focus on the patient, whilst that pertaining to gender tends toward demonstrating similar outcomes although there may be process variation. The study itself showed little difference in the process of reasoning between genders and ethnicities providing a degree of reassurance.

This potential issue was mitigated through providing all individuals with a summary of the narrative review and concept framework, in order to try and ensure a degree of homogeneity in understanding, and through the gathering of gender and ethnicity comparative data in the quantitative component of the study with a view to trying to establish some understanding of these issues.

The researcher was a member of the group and consequently there was a potential risk of them becoming dominant in the discussions negating the value of this process. To mitigate this one of the project supervisors acted as an independent facilitator to ensure that this was monitored and could be addressed if it occurred.

The question outputs represented the full spectrum of the model of reasoning produced, however there were only ten questions. The decision to use a limited number was to ensure that they were answered rapidly by the students following the relevant assessments and therefore hopefully represented their true thoughts. The number is low compared to tools such as the diagnostic thinking inventory and therefore it is possible that by only using small numbers of questions potentially important areas have not been fully covered or that the understanding gained will be limited. This is mitigated by the study design approach of mixed methods with the qualitative arm of the study designed to try and permit exploration

of these areas in more depth to avoid long questionnaires which risk students losing focus before the end.

11.4 The Mixed Methods Approach

11.4.1 Strengths

The choice of a mixed methods approach was in order to provide a more complete evaluation of the clinical reasoning process. The quantitative research element provides generalised information whilst the qualitative provides detail. The use of mixed methods has ensured that there is corroboration of the findings via the two underpinning methodologies as well as a depth of understanding of the more generalised results. This has ensured that there can be high confidence in the results. One single approach would be unlikely to have enabled such confidence in the results.

With respect to mixed methods more generally Creswell and Clark describe five key elements in a good mixed methods study (Creswell and Clark, 2017). These are shown in table 46.

Table 46 – The criteria for a good mixed methods study

1	Appropriate qualitative and quantitative data been collected in relation to the research question
2	The methods were pursued thoroughly and rigorously
3	The methods were integrated including representing the combined data appropriately
4	Appropriate mixed methods terms used within the study
5	The elements of the study fit together in a logical way

This study achieves all of the criteria set out in table 46 and therefore fulfils the criteria of a good mixed methods study.

Alternative approaches to mixed methods, such as talk aloud processes or microanalysis could not generate the same volume of data and would be subject to concerns as to whether students articulated what they truly did. Use of mixed methods in this way ensures that the two data sets can be used to complement one another and produce a result in which there can be confidence regarding the validity of the results.

11.4.2 Limitations

There are limitations to mixed methods research including complexity, ensuring that is the most appropriate way to research the question, and challenges of integration and interpretation. In order to mitigate these the study has taken each stage of the study and followed appropriate processes to maximise validity of the individual research components. The study also follows best practice mixed methods approaches to data integration and interpretation in order to provide reassurance on overall study validity.

11.5 Quantitative data Collection

11.5.1 Strengths

There are considerable strengths within the quantitative arm of the study. As stated in chapter 3 key considerations are sample size and the approach to statistical analysis in order to support validity. The sample size, whilst slightly below optimum, is large which serves to reduce the risk of sampling bias. Similarly the statistical approach, utilising parametric approaches with bootstrapping, showed results that were robust and can be considered valid. The decision to use quantitative approaches via Likert data items ensured a large amount of available data to which the qualitative data could then be used for complimentary consideration.

11.5.2 Limitations

The choice of sites for the study was designed to give a range of medical and physician associate students across different schools. Colleagues were contacted with a view to gaining a cross section of different medical, and physician associate, students. The sampling was convenience based in so far as those who were prepared to assist were included which led to varied numbers from the different schools with three schools involved. The spread means that there is medical student dominance from university 1 and physician associate dominance from university 3. The relative number from university 2 of both types means that meaningful quantitative statistical analysis between itself as a “new” medical school and University 1 as a long-established school is difficult to perform.

Importantly all three schools involved provide education based upon nationally set curricula for both medicine and physician associate studies and therefore it would be expected that students at any of them should be at a similar point educationally at the end of training. This

should mitigate the difference in student numbers from each of the different sites and permit a meaningful analysis.

Due to the need to manage the data collection within the processes set for the individual schools there was a difference in the exact structure of the OSCEs against which the students undertook the data collection. These are all described in full in chapter 6. Providing an identical OSCE experience was always going to be difficult due to the variation in different approaches at different schools in relation to station numbers, rest stations and physical space. The primary aim was that students completing the questions in the quantitative component of the study would have had suitable immersion in an OSCE environment to ensure the experience was at the forefront of their minds. In all cases the approach taken was able to ensure this with the formative OSCEs at university 1 and university 2, incorporating the test stations, whilst at university 3 the provision of the questions immediately after the summative OSCE ensured that this was the case. Similarly, the physical environment differed between the OSCEs (university 1 open lecture theatre, university 2 individual rooms or ward style, university 3 ward style) however all were appropriately constructed circuits for an OSCE at the required academic level. Whilst the exact OSCE structures were different all ensured appropriate immersion in the assessment modality immediately prior to data completion thus mitigating the potential limitations the variation might provide.

It proved extremely difficult to achieve the numbers considered optimum for the quantitative data based against the sample size calculations in chapter 5 (p78) for each individual element to be evaluated. Whilst theoretical numbers (based on individual cohort sizes) would have been sufficient this was a project based upon voluntary participation and thus those numbers could not be achieved. Whilst the total data points exceed the minimum these must be considered separately for SBA and OSCE as the majority of these represent one student providing two separate data points, one for each. In order to mitigate this issue the statistical analysis utilised parametric testing with bootstrapping acting as a sensitivity analysis. The similarity of the results demonstrate that the sample size is sufficient to allow inferences to be drawn appropriately.

11.6 Qualitative data collection

11.6.1 Strengths

The qualitative data collection was undertaken using best practice principles of qualitative research including the process of data capture and analysis. Thematic analysis, as described by Braun and Clarke, was used with each step followed appropriately. Focus groups were undertaken in suitable environments, with careful coding to seek reassurance regarding data saturation, ensuring robust complimentary data to support the mixed methods approach. The COREQ criteria have been considered and addressed with the checklist is included in this work as appendix 12 (Buus and Perron, 2020).

11.6.2 Limitations

Given the focus groups were not undertaken immediately after completing the questions recall bias becomes a potential risk. Previous researchers have looked at this and found that, in reality, the impact is relatively small (Neugebauer and Ng, 1990). The use of mixed methods for this study is designed to ensure that the qualitative data can be considered against the large volume of quantitative data thus seeking to reduce the potential impact of recall bias within the study as a whole.

Previous comment has been made on the focus group make up, especially away from university 1. All focus groups relied on volunteers. Away from university 1 this had to be opportunistic and therefore numbers were pragmatic, and all volunteers were welcome. Whilst the approach was similar at university 1 the proximity to the researcher led to an opportunity to better control the student numbers to an optimum point of 7-8 students per group however this remained challenging as shown by the group numbers. This has the potential limitation of leading to an imbalance in the development of subcategories. At university 2 the group was a mixed group of medical and physician associate students whilst at university 1 and university 3 the make-up was homogenous with either only medical, or physician associate, students. It was important to ensure that appropriate balance was gained between the two clinician types, so the six focus groups were made up of a mixed group at university 2, three medical student groups at university 1 and a physician associate group at each of university 1 and university 3. This gave a balanced mix of the two clinical groups seeking to address this potential limitation.

Whilst there was medical, and physician associate student, specific focus groups the data has been considered together which presents a potential limitation as the qualitative data cannot be considered side by side. As stated in the discussion (10.2.1 p191) the results suggest that the cognitive process in assessments is broadly identical between the two groups and so it would be expected that the qualitative output would also be similar mitigating this issue.

The decision on the number of focus groups was a pragmatic one based upon the targeted groups however there is always the concern in qualitative research as to exactly how many groups to utilise. Some authors recommend at least two groups per population group, in this case medical and physician associate students, giving a suggested minimum of four (two of each minimum) which was achieved in this work (Carey, 1995). The concept of saturation is based in grounded theory and not always applicable to studies which do not utilise this methodology with a previous systematic review finding that how saturation was judged was rarely stated (Carlsen and Glenton, 2011). One recent study has found that in general three to six groups will lead to over 90% saturation of all new data codes with the majority occurring in the first two (Hennink *et al.*, 2019). Following the approach of this study new codes added were recorded following the review of each focus group and shown in figure 11 (p138) in order to ensure that saturation was reached and provide reassurance around data validity thus addressing this potential limitation and ensuring transparency and rigor in approach.

11.7 Implicit versus Tacit Knowledge, its Implications

Within a discussion on the strengths and weaknesses of the study it is important to recognise the challenge that exists when researching a topic such as clinical reasoning. Whilst the knowledge required of clinicians is explicit (anatomy, pathology etc.) and can be easily shared, the process by which that knowledge is utilised in patient care is tacit, i.e. it exists within the head of the individual and can be difficult to articulate and specify. A strength of this study is its attempt to understand this tacit process. It has been noted that there is a clear epistemological dichotomy in the assumption that expertise is important whilst experience is untrustworthy in the clinical reasoning process (Tonelli and Shapiro, 2020). This study has considerable strengths in seeking to understand this tacit process through the use of mixed methods, but it remains limited by the ability of students to

adequately express their reasoning process which, as discussed previously, is multifaceted and task specific. Authors have specifically noted that effective clinical education relies upon the inclusion of elements of tacit knowledge yet it remains poorly covered in curricula (Heiberg Engel, 2008).

Whilst this research is situated within one specific area, assessment, we still cannot, with complete certainty, conclude that the concept framework is based upon explicit truth. Recognising that the reasoning process develops from novice to expert through experience we must note that these students are relatively early in their development of this process. Whilst the process followed has many strengths this limitation must be acknowledged within this work.

11.8 The Researcher as an Insider

The researcher has a relationship to the participants that can be described as endogenous, meaning they are classed as an insider researcher. An insider researcher undertakes the research within a community or group with which they are also a member, or have significant and intimate knowledge of (Brannick and Coghlan, 2007). The researcher is a doctor with significant knowledge and expertise in the areas under research as well as the target groups, working as course lead for the university 1 physician associate course, involved with the medicine course and having supported multiple SBA and OSCE examinations. This provides significant strength to this study, particularly in relation to data interpretation, though the risk of researcher bias in the delivery of the qualitative component exists.

11.8.1 Power dynamics in the recruitment

This research involved power dynamics between the researcher and the participants with the potential to affect the way that groups were constructed and risks of coercion in recruitment (Fleming, 2018). The least at risk of this was at university 2 with which there was no prior engagement with the participants. The physician associate course at university 3 had a slightly different power dynamic in so far as the researcher is external examiner for the course, but this involves minimal contact with the students prior to this research and, again, recruitment was voluntary and facilitated by a member of the academic staff there. University 1 had the greatest risk of impact from the power dynamic. The researcher has

limited day to day involvement with the medicine course and recruitment was voluntary and managed via the student medical society thus removing the researcher from direct recruitment contact. The greatest risk was with the physician associate course where the researcher themselves requested volunteers. In this case no incentive was offered nor direct request made, other than to email prospective participants, to reduce any feeling of mandatory participation by students. For this final group it was also ensured that the researcher was not involved as an examiner in their final summative OSCE, whilst overseeing its conduct, to ensure that there was no proxy concern that non-participation may be perceived as having a potential consequence risk.

In all of the above the use of non-researcher recruiters (where possible), or non-direct contact via email, sought to mitigate the power dynamic in recruitment to the study and therefore contribute to reducing the risk of bias, or accusations thereof.

11.8.2 Insider risks to qualitative data interpretation

Insider researchers have significant advantages in data interpretation, in so far as the language and nature of the topic is well known to them, preventing the need to develop this understanding and reducing the likelihood that the responses in qualitative interviews are misunderstood (Brannick and Coghlan, 2007). This is a double-edged sword however as the risk is that the researcher makes assumptions regarding the expected patterns in the data and thus fails to address unexpected findings or address incongruous data (Mercer, 2007).

Addressing this potential limitation is about recognising this, openly acknowledging it, and ensuring accuracy in the coding along with ensuring appropriate external scrutiny of the data outputs relative to content and welcoming challenge in the data interpretation process.

11.9 Summary

This chapter has set out the strengths and potential limitations, that exist in this work broken down by section. In the case of the limitations the mitigations and approach to ensure appropriate rigor have been stated. By considering the strengths, and actively seeking out and acknowledging potential limitations, the process can be considered for rigor and appropriate consideration regarding potential mitigations can be ensured to have been taken.

Chapter 12 – Conclusion

12.1 Introduction

This work set out to answer the following question: How do final year clinical students undertake the process of clinical reasoning in SBA and OSCE assessments? This conclusion will consider the research question before briefly summarising the implications for policy, education and research.

I will reflect as to the extent to which I have answered the research question and what the study findings suggest for future research avenues extending from this work.

12.2 Considering the research question

12.2.1 Answering the research question

The research set out to answer the question as to how final year clinical students undertake the process of clinical reasoning in SBA and OSCE assessments. In this respect the study was successful. The study is distinctive as there are no other studies looking directly at this question, nor that have done so through the use of multiple techniques (literature review to create a unique concept framework, use of a consensus group to develop questions and then a mixed methods study approach).

The results show that clinical students have a high preference for inductive approaches in the studied assessments, especially in SBA examinations, as their primary approach to conducting clinical reasoning. Early hypothesis generation is typical in both assessment types. Satisficing is common, although differs slightly in rationale, in the two assessment types. Finally students seem to have a belief that they understand biases, but the evidence suggests that this is not the case.

The finding that clinical students have a high preference for inductive approaches in clinical reasoning was evident through the high impact of demographic details in decision making and, to a lesser extent, the clinical setting of the question.

Early hypothesis generation is typical within the approach taken. In SBA style examinations the early hypothesis is commonly explored in parallel with the answer formulation on the basis that “what you see is all there is” whereas students in OSCE use the early hypothesis

as a key part of the process to enable the decisions around how to structure their approach to the station. This forms part of an abductive clinical reasoning process, albeit one in which the primary clinical reasoning approach chosen is induction. Reflecting this to the process of abduction, the formation of an early hypothesis constitutes the perception and imagination components of the process as initial impressions of the assessment data (perception) lead to the individual to formulate the hypothesis (imagination and creativity). This hypothesis is then explored through a chosen clinical reasoning approach to formulate an answer. When students reason in SBA assessments the abductive process is more compact and in doing so diverts away from the reality of clinical reasoning. A more real-life clinical reasoning approach is seen in the OSCE examination where there is more of an exploration of the clinical scenario, however the nature of assessment is such that a true diagnostic endpoint, whilst desirable, may not be required to do well in the marking process and thus this potentially impacts the clinical reasoning undertaken.

Regarding the closure of the clinical reasoning process satisficing is evident, particularly where individual knowledge is lacking, or time pressure leads to students having to make decisions quickly in SBAs. It was also evident in the overall approach to OSCE examinations where the focus was on acquisition of points rather than the diagnostic endpoint as highlighted above. Bias recognition is shown to be less well understood. Whilst students talked about biases it was in relation to aspects of pattern recognition in many cases. There appeared to be confusion between the concepts of biases, as a negative cognitive effect, and heuristics, as positive decisions to short cut the clinical reasoning process based on particular characteristics within the decision itself. The most obvious example of heuristics in relation to assessment was that of ethnicity driving decision making on likely illness, specifically that its inclusion must lead to an answer that is directly related to the ethnicity itself.

The initial part of this study involved developing a unique model of clinical reasoning from literature representing a range of clinical professions. Chapters nine and ten presented and discussed the results and facilitated the development of a revised concept framework (Figure 17, page 193). This framework can be used to help develop appropriate schemes of assessment and support the development of pedagogical approaches to clinical reasoning.

This study has allowed us to understand how students are undertaking the process of clinical reasoning however it also highlights the difficulties in making the assessments "real world" appropriate with respect to the clinical reasoning processes we seek to achieve from our students.

12.2.2 The implications of these findings

These findings are relevant in that they give us a clear picture as to how students undertake the process of clinical reasoning in these assessment types and thus permit us to reflect upon the way that the primary assessment types used in clinical education lead to the utilisation of specific approaches in students. This is important because, if we recognise that our role as educators is to develop a wide range of clinical reasoning approaches in our students, we need to consider our overall scheme of assessment based upon this knowledge, as well as reflecting upon how our pedagogical approaches lead to the development of specific patterns of thinking.

With respect to policy, it is important that clinical reasoning, both diagnostic and therapeutic, is seen as being just as vital as other aspects of medical education such as biomedical knowledge. There is limited value in knowing the what and why of clinical care if we fail to teach the how. The bodies responsible for the medical curriculum, such as the General Medical Council, need to embrace this issue and scrutinise it in their curriculum development, and course quality assurance processes, to focus course leaders on the way it is transacted. With the evidence of clinical reasoning errors directly leading to poor patient outcomes instilling a requirement for this in policy is crucial.

For education there are clear implications in the way that we teach and assess clinical reasoning which will require a shift in the way that curricula are planned, delivered and assessed. There are clear implications for faculty training that will need careful consideration, particularly in the workplace, as teaching will often be delivered by more junior clinicians who are not the usual target for these developments. All those involved in teaching our students need to better understand their responsibilities with respect to developing clinical reasoning in students. This work is also significant because it challenges our assumptions that our assessments are effective in ensuring that students are appropriately prepared for clinical reasoning in practice. This means that we are required to

reflect upon our scheme of assessment and its relevance for purpose, namely appropriate preparation of “day one” practitioners. Programmatic assessment must become the norm with consideration given to an appropriate variety of assessments to ensure that clinical reasoning is appropriately embedded.

When considering implications for research there is evidence that aspects of clinical reasoning need further consideration. Whilst other assessment methods, such as workplace-based assessment, require evaluation for us to understand if they assess different aspects of reasoning, there also needs to be more research into full programmes of assessment and how they provide students with insight into multiple clinical reasoning approaches. Research is also needed as to how we deliver effective education in clinical reasoning, both in the interaction between educators and students, but also how faculty are supported in their wider understanding of the subject and delivery to students in the differing pedagogical environments.

12.2.3 Applicability to assessment and teaching in real world settings

The previous section has commented upon the implications for education, policy and research which includes particular comments with respect to education in the clinical setting. Whilst the conclusions that can be immediately drawn are limited due to the study parameters the concept framework developed can be used to help explain clinical reasoning within real world settings. Its usage, along with micro analytic approaches, based around its components can facilitate conversations on clinical reasoning and its application in practice. The study findings should also act as a trigger for us to ask questions around the clinical reasoning process being followed when students are being taught through the analysis of real patient cases and to allow us to challenge, support and develop their approaches.

12.3 Research limitations and suggestions for future research

12.3.1 Generalisability

It is important to acknowledge that this research has limitations. The study looked specifically at the process of clinical reasoning within two specific assessment types rather than more broadly and also does not explore the wider curriculum and how aspects of it will contribute to the student’s development of this cognitive process. This is pragmatic in order to ensure focus to the work, however it also limits the conclusions that can be drawn. The

work has allowed the development of a concept framework for clinical reasoning however this can only be said to be proven as applicable to the two methods of assessment studied. To validate this concept framework in other scenarios further research needs to be undertaken evaluating it against other assessment types and "real world" clinical interactions in order to confirm that it stands true to all situations. Even then we will have to recognise all of the potential limitations that have previously been discussed (do people do what they say they do), albeit that this has already been shown to be a limitation in all research on the clinical reasoning process.

12.3.2 The type of assessments studied

The study only looked at the two primary types of assessment undertaken within clinical courses and did not look at workplace-based assessment. Workplace-based assessment forms another key part of our overall scheme of assessment and this has not been covered within this study other than a consideration within the literature review and in the discussion chapter. This raised some key issues, in particular around the training of assessors, which is backed up through work showing that students themselves had concerns regarding the subjectivity of such assessments, and the attitudes of assessors, diminishing its value (Nesbitt *et al.*, 2013). Further research is needed to understand the process of clinical reasoning that students are undertaking in workplace-based assessments. This needs to include the suitability of these assessments to support appropriate clinical reasoning development, their alignment to how we develop the student's clinical reasoning approaches through the curriculum and how we train assessors to recognise and support this process to complement reasoning development.

12.3.3 The clinical roles studied

This study looked at medical students and student physician associates. This was due to both having a medical model of clinical training with an expectation that this would lead to a similar pattern of clinical reasoning. This appears to be the case however it does mean that, whilst the findings can be applied to clinicians with this style of education, we cannot be certain that the concept framework is applicable across all clinicians, albeit that it was initially developed from the multi-professional literature.

Further research is required to establish if the work is generalisable for all clinician types both with respect to the clinical reasoning processes within assessments but also to the relevance of the concept framework. If it holds true across other professional groups, then it has the potential to support effective inter-professional education on the topic of clinical reasoning.

12.4 To Conclude

In conclusion this study has explored the research question, specifically as to how students undertake the process of clinical reasoning in assessments, through a mixed methods approach. This utilised a narrative review to consider the literature that exists on the topic to produce a concept framework which was then utilised in a nominal group meeting to set the questions for use in the two arms of the study. The study evaluated the data using methodology appropriate to the two data types before considering them together in a convergent approach. The narrative conclusion reached regarding the approach permitted a revision of the concept framework and the implications for assessment and education to be considered.

The hypothesis that students would show a mixed pattern of clinical reasoning, employing both inductive and deductive methods, when answering questions in examinations has not been shown to hold entirely true with a strong preference toward induction.

It is hoped that this work can lead to further research on the topic of clinical reasoning, both in assessment and education, supported by the concept framework developed in this work and its inevitable later iterations.

References

Ahlsen, B., Mengshoel, A. M., Bondevik, H. and Engebretsen, E. (2018) 'Physiotherapists as detectives: investigating clues and plots in the clinical encounter', *Medical Humanities*, 44(1), pp. 40-45.

Allen, I. E. and Seaman, C. A. (2007) 'Likert scales and data analyses', *Quality progress*, 40(7), p. 64.

Amin, M. and Wang, J. (2009) 'Routine laboratory testing to evaluate for medical illness in psychiatric patients in the emergency department is largely unrevealing', *The Western Journal of Emergency Medicine*, 10(2), pp. 97-100.

Anderson, C. (2010) 'Presenting and evaluating qualitative research', *American Journal of Pharmaceutical Education*, 74(8), pp. 141-141.

Andrews, J. J. W. and Syeda, M. M. (2017) 'Clinical Reasoning in School Psychology: From Assessment to Intervention', *Canadian Journal of School Psychology*, 32(1), pp. 3-15.

Anonymous (2014) 'Kick the bar chart habit', *Nature methods*, 11(2), pp. 113-113.

Aristotle (1926) *Nicomachean ethics*. revised. edn. Edited by Rackham, H. Cambridge, MA: Cambridge, MA : Harvard University Press, 1926.

Ark, T. K., Brooks, L. R. and Eva, K. W. (2006) 'Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices?', *Academic Medicine*, 81(4), pp. 405-409.

Artino, A. R. and Jones, K. D. (2013) 'AM Last Page: Self-Regulated Learning—A Dynamic, Cyclical Perspective', *Academic Medicine*, 88(7), pp. 1048-1048.

Atieno, O. P. (2009) 'An analysis of the strengths and limitation of qualitative and quantitative research paradigms', *Problems of Education in the 21st Century*, 13, pp. 13-18.

Baethge, C., Goldbeck-Wood, S. and Mertens, S. (2019) 'SANRA-a scale for the quality assessment of narrative review articles', *Research Integrity and Peer Review*, 4(1), pp. 5-5.

Baid, H. (2006) 'Differential diagnosis in advanced nursing practice', *British Journal of Nursing*, 15(18), pp. 1007-1011.

Baker, E. A., Ledford, C. H., Fogg, L., Way, D. P. and Park, Y. S. (2015) 'The IDEA Assessment Tool: Assessing the Reporting, Diagnostic Reasoning, and Decision-Making Skills Demonstrated in Medical Students' Hospital Admission Notes', *Teaching and Learning in Medicine*, 27(2), pp. 163-173.

Balla, J. I., Heneghan, C., Glasziou, P., Thompson, M. and Balla, M. E. (2009) 'A model for reflection for good clinical practice', *Journal of Evaluation in Clinical Practice*, 15(6), pp. 964-969.

Barrett, J. L., Denegar, C. R. and Mazerolle, S. M. (2018) 'Challenges Facing New Educators: Expanding Teaching Strategies for Clinical Reasoning and Evidence-Based Medicine', *Athletic Training Education Journal*, 13(4), pp. 359-366.

Ben-David, M. F. (2000) 'AMEE Guide No. 18: Standard setting in student assessment', *Medical Teacher*, 22(2), pp. 120-130.

Biggs, J. (1996) 'Enhancing Teaching through Constructive Alignment', *Higher Education*, 32(3), pp. 347-364.

Black, T. R. (1999) *Doing quantitative research in the social sciences : an integrated approach to research design, measurement and statistics*. London: London : SAGE, 1999.

Brannick, T. and Coghlan, D. (2007) 'In Defense of Being "Native": The Case for Insider Academic Research', *Organizational Research Methods*, 10(1), pp. 59-74.

Braude, H. D. (2009) 'Clinical intuition versus statistics: different modes of tacit knowledge in clinical epidemiology and evidence-based medicine', *Theoretical Medicine and Bioethics*, 30(3), pp. 181-198.

Braun, V. and Clarke, V. (2013) *Successful qualitative research : a practical guide for beginners*. Edited by Clarke, V. London: London : Sage Publications, 2013.

Braun, V. and Clarke, V. (2021) 'One size fits all? What counts as quality practice in (reflexive) thematic analysis?', *Qualitative Research in Psychology*, 18(3), pp. 328-352.

Brown, C., Goss, C. and Sam, A. H. (2023) 'Is the awarding gap at UK medical schools influenced by ethnicity and medical school attended? A retrospective cohort study', *BMJ Open*, 13(12), pp. e075945-e075945.

Bryman, A. (2006) 'Integrating quantitative and qualitative research: how is it done?', *Qualitative Research : QR*, 6(1), pp. 97-113.

Buus, N. and Perron, A. (2020) 'The quality of quality criteria: Replicating the development of the Consolidated Criteria for Reporting Qualitative Research (COREQ)', *International journal of nursing studies*, 102, pp. 103452-103452.

Carey, M. A. (1995) 'Comment: Concerns in the Analysis of Focus Group Data', *Qualitative Health Research*, 5(4), pp. 487-495.

Carlsen, B. and Glenton, C. (2011) 'What about N? A methodological study of sample-size reporting in focus group studies', *BMC Medical Research Methodology*, 11(1), pp. 26-26.

Cassady, J. C. and Johnson, R. E. (2002) 'Cognitive Test Anxiety and Academic Performance', *Contemporary Educational Psychology*, 27(2), pp. 270-295.

Ceolin, S., González, J. S., Ruiz, M. C. S. and Heck, R. M. (2017) 'Theoretical bases of critical thinking in ibero-american nursing: Integrative literature review', *Texto e Contexto Enfermagem*, 26(4).

Cerdeña, J. P., Jaswaney, R. and Plaisime, M. V. (2022) 'Race, Ethnicity, and Immigration Status in a Medical Licensing Educational Resource: a Systematic, Mixed-Methods Analysis', *Journal of General Internal Medicine*, 37(5), pp. 1045-1051.

Çetinkaya, L., Keser, I. L., Yildirim, S. and Keser, H. (2024) 'The effect of case-based mobile virtual patient application on students' academic achievement in clinical reasoning skills', *Medical Education Online*, 29(1), pp. 2322223-2322223.

Chaffey, L., Unsworth, C. and Fossey, E. (2010) 'A grounded theory of intuition among occupational therapists in mental health practice', *British Journal of Occupational Therapy*, 73(7), pp. 300-308.

Charlin, B., Tardif, J. and Boshuizen, H. P. A. (2000) 'Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research', *Academic Medicine*, 75(2), pp. 182-190.

Christensen, J. H. (2022) 'Enhancing mixed methods pragmatism with systems theory: Perspectives from educational research', *Systems research and behavioral science*, 39(1), pp. 104-115.

Clarke, V. and Braun, V. (2013) *Successful qualitative research : a practical guide for beginners*. Edited by Braun, V. London: London : SAGE, 2013.

Cleary, T. J., Durning, S. J. and Artino, A. R. (2016) 'Microanalytic Assessment of Self-Regulated Learning During Clinical Reasoning Tasks: Recent Developments and Next Steps', *Academic Medicine*, 91(11), pp. 1516-1521.

Cohen, L. (2018) *Research methods in education*. Eighth edition / Louis Cohen, Lawrence Manion and Keith Morrison. edn. Edited by Manion, L. and Morrison, K. London, New York: London, New York : Routledge, Taylor & Francis Group, 2018.

Cooke, A., Smith, D. and Booth, A. (2012) 'Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis', *Qualitative Health Research*, 22(10), pp. 1435-1443.

Cox, A. and Li, S. (2020) 'The Medical Consultation through the Lenses of Language and Social Interaction Theory', *Advances in Health Sciences Education*, 25(1), pp. 241-257.

Creswell and Clark, V. L. P. (2017) *Designing and conducting mixed methods research*. 3rd edn. Los Angeles: Sage publications.

Creswell, J. W., Fetters, M. D. and Ivankova, N. V. (2004) 'Designing a mixed methods study in primary care', *Annals of Family Medicine*, 2(1), pp. 7-12.

Croskerry, P. (2009a) 'Context Is Everything or How Could I Have Been That Stupid?', *Healthcare Quarterly*, 12(Sp), pp. e171-e176.

Croskerry, P. (2009b) 'A Universal Model of Diagnostic Reasoning', *Academic Medicine*, 84(8), pp. 1022-1028.

Crossley, J., Johnson, G., Booth, J. and Wade, W. (2011) 'Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales', *Medical Education*, 45(6), pp. 560-569.

Custers, E. J. F. M. (2013) 'Medical education and cognitive continuum theory: an alternative perspective on medical problem solving and clinical reasoning', *Academic Medicine*, 88(8), pp. 1074-1080.

Custers, E. J. F. M. (2019) 'Theories of truth and teaching clinical reasoning and problem solving', *Advances in Health Sciences Education*, 24(4), pp. 839-848.

Custers, E. J. F. M., Boshuizen, H. P. A. and Schmidt, H. G. (1998) 'The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results From an Interview Study', *Cognition and Instruction*, 16(4), pp. 367-398.

Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R., Sergio Da Silva, A., Cleary, T., Stojan, J. and Gruppen, L. D. (2019) 'Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance', *Academic Medicine*, 94(6), pp. 902-912.

de Winter, J. C. F. and Dodou, D. (2010) 'Five-point likert items: T test versus Mann-Whitney-Wilcoxon', *Practical assessment, research & evaluation*, 15(11), p. 11.

Delany, C. and Watkin, D. (2009) 'A Study of Critical Reflection in Health Professional Education: "Learning Where Others Are Coming from"', *Advances in Health Sciences Education*, 14(3), pp. 411-429.

Denzin, N. K. (1989) *The research act : a theoretical introduction to sociological methods*. Englewood Cliffs, N.J.: Prentice Hall.

Denzin, N. K. and Lincoln, Y. S. (2011) *The Sage handbook of qualitative research*. 4th ed. edn. Thousand Oaks: Thousand Oaks : Sage, c2011.

Descartes, R. (1649) *A discourse of a method for the well-guiding of reason, and the discovery of truth in the sciences. [electronic resource]*. London: London : Printed by Thomas Newcombe, for John Holden at the Anchour in the new Exchange, 1649.

Durning, S. J., Artino, A., Boulet, J., La Rochelle, J., Van Der Vleuten, C., Arze, B. and Schuwirth, L. (2012a) 'The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning', *Medical Teacher*, 34(1), pp. 30-37.

Durning, S. J. and Artino, A. R. (2011) 'Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52', *Medical Teacher*, 33(3), pp. 188-199.

Durning, S. J., Artino, A. R., Boulet, J. R., Dorrance, K., van der Vleuten, C. and Schuwirth, L. (2012b) 'The Impact of Selected Contextual Factors on Experts' Clinical Reasoning Performance (Does Context Impact Clinical Reasoning Performance in Experts?)', *Advances in Health Sciences Education*, 17(1), pp. 65-79.

Durning, S. J., Artino, A. R., Jr., Schuwirth, L. and van der Vleuten, C. (2013) 'Clarifying Assumptions to Enhance Our Understanding and Assessment of Clinical Reasoning', *Academic Medicine*, 88(4), pp. 442-448.

Dzau, V. J., Kirch, D. G. and Nasca, T. J. (2018) 'To Care Is Human — Collectively Confronting the Clinician-Burnout Crisis', *New England Journal of Medicine*, 378(4), pp. 312-314.

Elliott RA, Camacho E, Jankovic D*, et al*. (2021) 'Economic analysis of the prevalence and clinical and economic burden of medication error in England', *BMJ Quality & Safety*, 30**,** pp. 96-105.

Elstein, A. S., Shulman, L. S. and Sprafka, S. A. (1990) 'Medical Problem Solving: A Ten-Year Retrospective', *Evaluation & the Health Professions*, 13(1), pp. 5-36.

Enns, M. W., Cox, B. J., Sareen, J. and Freeman, P. (2001) 'Adaptive and maladaptive perfectionism in medical students: A longitudinal investigation', *Medical Education*, 35(11), pp. 1034-1042.

Epstein, R. M. and Hundert, E. M. (2002) 'Defining and assessing professional competence', *Journal of the American Medical Association*, 287(2), pp. 226-235.

Ercikan, K. and Roth, W.-M. (2006) 'What Good Is Polarizing Research into Qualitative and Quantitative?', *Educational Researcher*, 35(5), pp. 14-23.

Ericsson, K. A. (2007) 'An expert-performance perspective of research on medical expertise: the study of clinical performance', *Medical Education*, 41(12), pp. 1124-1130.

Esteves, J. E., Bennison, M. and Thomson, O. P. (2013) 'Script concordance test: Insights from the literature and early stages of its implementation in osteopathy', *International Journal of Osteopathic Medicine*, 16(4), pp. 231-239.

Eva, K. W. (2005) 'What every teacher needs to know about clinical reasoning', *Medical Education*, 39(1), pp. 98-106.

Evans, J. and Benefield, P. (2001) 'Systematic Reviews of Educational Research: Does the medical model fit?', *British Educational Research Journal*, 27(5), pp. 527-541.

Fabre, V., Davis, A., Diekema, D. J., Granwehr, B., Hayden, M. K., Lowe, C. F., Pfeiffer, C. D., Sick-Samuels, A. C., Sullivan, K. V., Van Schooneveld, T. C. and Morgan, D. J. (2023) 'Principles of diagnostic stewardship: A practical guide from the Society for Healthcare Epidemiology of America Diagnostic Stewardship Task Force', *Infection Control and Hospital Epidemiology*, 44(2), pp. 178-185.

Feufel, M. A. and Flach, J. M. (2019) 'Medical education should teach heuristics rather than train them away', *Medical Education*, 53(4), pp. 334-344.

Fleming, J. (2018) 'Recognizing and resolving the challenges of being an insider researcher in work-integrated learning', *International Journal of Work-integrated Learning*, 19(3), pp. 311-320.

Flick, U. (2017) 'Mantras and Myths: The Disenchantment of Mixed-Methods Research and Revisiting Triangulation as a Perspective', *Qualitative Inquiry*, 23(1), pp. 46-57.

Flick, U. (2018) *An Introduction to Qualitative Research*. 6th edn. London: SAGE.

Geisler, P. R. and Lazenby, T. W. (2009) 'Clinical Reasoning in Athletic Training Education: Modeling Expert Thinking', *Athletic Training Education Journal*, 4(2), pp. 52-65.

Ghadiri, S. J. (2020) 'Physician associates: an asset for physician training and a 21st-century NHS?', *Future Healthcare Journal*, 7(3), pp. e9-e10.

Gibson, D., Velde, B., Hoff, T., Kvashay, D., Manross, P. L. and Moreau, V. (2000) 'Clinical Reasoning of a Novice versus an Experienced Occupational Therapist: A Qualitative Study', *Occupational Therapy in Health Care*, 12(4), pp. 15-31.

Gigerenzer, G. and Goldstein, D. G. (1996) 'Reasoning the Fast and Frugal Way: Models of Bounded Rationality', *Psychological Review*, 103(4), pp. 650-669.

Gingerich, A., Regehr, G. and Eva, K. W. (2011) 'Rater-based assessments as social judgments: rethinking the etiology of rater errors', *Academic Medicine*, 86(10), pp. S1-7.

Gishen, F. and Lokugamage, A. (2019) 'Diversifying the medical curriculum', *British Medical Journal*, 364, p. l300.

Glaser, B. G. (1967) *The discovery of grounded theory : strategies for qualitative research*. Edited by Strauss, A. L. New York: New York : Aldine de Gruyter, 1967.

GMC, T. (2017) *Generic professional capabilities framework*. Available at: www.gmc-uk.org/-/media/documents/generic-professional-capabilities-framework_pdf-108338152.pdf (Accessed: 11th January 2025).

GMC, T. (2020) *Outcomes for Graduates*. Available at: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/outcomes-for-graduates/outcomes-for-graduates (Accessed: 11th January 2025).

GMC, T. (2024) *General Medical Council Website*. Available at: https://www.gmc-uk.org/education/medical-licensing-assessment (Accessed: 31st May 2024).

Goldowsky, A. and Rencic, J. (2023) 'Self-regulated learning and the future of diagnostic reasoning education', *Diagnosis (Berlin, Germany)*, 10(1), pp. 24-30.

Gonullu, I., Bayazit, A. and Erden, S. (2024) 'Exploring medical students' perceptions of individual and group-based clinical reasoning with virtual patients: a qualitative study', *BMC Medical Education*, 24(1), pp. 189-189.

Google (2022a) *Google Meet* (Version Meet PWA) [Computer program]. (Accessed: 22nd November 2022)

Google (2022b) *Google Sheets* (Version Sheets CWP) [Computer program]. (Accessed: 22nd November 2022)

Gopal, D. P., Chetty, U., O'Donnell, P., Gajria, C. and Blackadder-Weinstein, J. (2021) 'Implicit bias in healthcare: clinical practice, research and decision making', *Future Healthcare Journal*, 8(1), pp. 40-48.

Gormley, G. J., Johnston, J. L., Cullen, K. M. and Corrigan, M. (2021) 'Scenes, symbols and social roles: raising the curtain on OSCE performances', *Perspectives in Medical Education*, 10(1), pp. 14-22.

Graber (2020) 'Progress understanding diagnosis and diagnostic errors: thoughts at year 10', *Diagnosis*, 7(3), pp. 151-159.

Graber, Rencic, J., Rusz, D., Papa, F., Croskerry, P., Zierler, B., Harkless, G., Giuliano, M., Schoenbaum, S., Colford, C., Cahill, M. and Olson, A. P. J. (2018) 'Improving diagnosis by improving education: a policy brief on education in healthcare professions', *Diagnosis*, 5(3), pp. 107-118.

Graber, M., Gordon, R. and Franklin, N. (2002) 'Reducing diagnostic errors in medicine: What's the goal?', *Academic Medicine*, 77(10), pp. 981-992.

Greene, J. C., Caracelli, V. J. and Graham, W. F. (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis*, 11(3), pp. 255-274.

Greenhalgh, T. (1999) 'Narrative based medicine: narrative based medicine in an evidence based world', *British Medical Journal*, 318(7179), pp. 323-5.

Gregory, A. T. and Denniss, A. R. (2018) 'An Introduction to Writing Narrative and Systematic Reviews — Tasks, Tips and Traps for Aspiring Authors', *Heart, Lung & Circulation*, 27(7), pp. 893-898.

Gruppen, L. D., ten Cate, O., Lingard, L. A., Teunissen, P. W. and Kogan, J. R. (2018) 'Enhanced Requirements for Assessment in a Competency-Based, Time-Variable Medical Education System', *Academic Medicine*, 93(3), pp. S17-S21.

Gruppen, L. D., Woolliscroft, J. O. and Wolf, F. M. (1988) 'The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses', *Research in Medical Education: Proceedings of the Annual Conference. Conference on Research in Medical Education*, 27, pp. 242-7.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P. and Schünemann, H. J. (2008) 'GRADE: an emerging consensus on rating quality of evidence and strength of recommendations', *British Medical Journal*, 336(7650), pp. 924-926.

Hall, P. and Hart, J. D. (1990) 'Bootstrap Test for Difference between Means in Nonparametric Regression', *Journal of the American Statistical Association*, 85(412), pp. 1039-1049.

Hamm, R. M. (2004) 'Theory about Heuristic Strategies Based on Verbal Protocol Analysis: The Emperor Needs a Shave', *Medical Decision Making*, 24(6), pp. 681-686.

Hampton, J. R., Harrison, M. J., Mitchell, J. R., Prichard, J. S. and Seymour, C. (1975) 'Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients', *British Medical Journal*, 2(5969), pp. 486-489.

Heiberg Engel, P. J. (2008) 'Tacit knowledge and visual expertise in medical diagnostic reasoning: Implications for medical education', *Medical teacher*, 30(7), pp. e184-e188.

Hennink, M. M., Kaiser, B. N. and Weber, M. B. (2019) 'What Influences Saturation? Estimating Sample Sizes in Focus Group Research', *Qualitative Health Research*, 29(10), pp. 1483-1496.

Hesse-Biber, S. (2010) 'Qualitative Approaches to Mixed Methods Practice', *Qualitative Inquiry*, 16(6), pp. 455-468.

Higgs, J., Burn, A. and Jones, M. (2001) 'Integrating clinical reasoning and evidence-based practice', *AACN clinical issues*, 12(4), pp. 482-490.

Hodges, B. and McIlroy, J. H. (2003) 'Analytic global OSCE ratings are sensitive to level of training', *Medical Education*, 37(11), pp. 1012-1016.

Hogg, W., Baskerville, N. and Lemelin, J. (2005) 'Cost savings associated with improving appropriate and reducing inappropriate preventive care: Cost-consequences analysis', *BMC Health Services Research*, 5(1), pp. 20-20.

Holder, A. G. (2018) 'Clinical Reasoning: A State of the Science Report', *International journal of Nursing Education Scholarship*, 15(1).

Horst, W. J. R. and Webber, M. M. (1973) 'Dilemmas in a General Theory of Planning', *Policy Sciences*, 4(2), pp. 155-169.

Hrynchak, P., Glover Takahashi, S. and Nayer, M. (2014) 'Key-feature questions for assessment of clinical reasoning: a literature review', *Medical Education*, 48(9), pp. 870-883.

Humphrey-Murto, S., Varpio, L., Gonsalves, C. and Wood, T. J. (2017a) 'Using consensus group methods such as Delphi and Nominal Group in medical education research', *Medical Teacher*, 39(1), pp. 14-19.

Humphrey-Murto, S., Varpio, L., Wood, T. J., Gonsalves, C., Ufholz, L.-A., Mascioli, K., Wang, C. and Foth, T. (2017b) 'The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review', *Academic Medicine*, 92(10), pp. 1491-1498.

Hunter, K. (1996) '"Don't think zebras": Uncertainty, interpretation, and the place of paradox in clinical education', *Theoretical Medicine and Bioethics*, 17(3), pp. 225-241.

Huwendiek, S., Reichert, F., Duncker, C., de Leng, B. A., van der Vleuten, C. P. M., Muijtjens, A. M. M., Bosse, H.-M., Haag, M., Hoffmann, G. F., Tönshoff, B. and Dolmans, D. (2017) 'Electronic assessment of clinical reasoning in clerkships: A mixed-methods comparison of long-menu key-feature problems with context-rich single best answer questions', *Medical Teacher*, 39(5), pp. 476-485.

Hydén, L. C. and Bülow, P. H. (2003) 'Who's talking: drawing conclusions from focus groups-some methodological considerations', *International Journal of Social Research Methodology*, 6(4), pp. 305-321.

IBM (2021) *SPSS Statistics for Windows* (Version 28.0) [Computer program]. IBM Corp. (Accessed: 22nd November 2022)

Ilgen, J. S., Humbert, A. J., Kuhn, G., Hansen, M. L., Norman, G. R., Eva, K. W., Charlin, B., Sherbino, J. and Kowalenko, T. (2012) 'Assessing Diagnostic Reasoning: A Consensus Statement Summarizing Theory, Practice, and Future Needs', *Academic Emergency Medicine*, 19(12), pp. 1454-1461.

Jackson, B., Marshall, M. and Schofield, S. (2017) 'Barriers and facilitators to integration of physician associates into the general practice workforce: A grounded theory approach', *British Journal of General Practice*, 67(664), pp. e785-e791.

James, K., Cross, K., Lucarotti, M. E., Fowler, A. L. and Cook, T. A. (2009) 'Undertaking procedure-based assessment is feasible in clinical practice', *Annals of the Royal College of Surgeons of England*, 91(2), pp. 110-112.

Jay, R., Sandars, J., Patel, R., Leonardi-Bee, J., Ackbarally, Y., Bandyopadhyay, S., Faraj, D., O'Hanlon, M., Brown, J. and Wilson, E. (2024) 'The Use of Virtual Patients to Provide Feedback on Clinical Reasoning: A Systematic Review', *Academic Medicine*, 100(2), p229-238

Jensen, J. L., Berry, D. A. and Kummer, T. A. (2013) 'Investigating the Effects of Exam Length on Performance and Cognitive Fatigue', *PloS one*, 8(8), pp. e70270-e70270.

John, J. N., Linda, L. B., Duffy, F. D. and Gregory, S. F. (2003) 'The Mini-CEX: A Method for Assessing Clinical Skills', *Annals of Internal Medicine*, 138(6), pp. 476-481.

Johnson-Laird, P. a. L., N.L. (2006) 'Are There Cross-Cultural Differences in Reasoning?', *Proceedings of the Annual Meeting of the Cognitive Science Society*, 28, pp. 459-464.

Jones, J. and Hunter, D. (1995) 'Consensus Methods For Medical And Health Services Research', *British Medical Journal*, 311(7001), pp. 376-380.

Jones, S. R., Carley, S. and Harrison, M. (2003) 'An introduction to power and sample size estimation', *Emergency Medicine Journal*, 20(5), pp. 453-458.

Kahlke, R. and Eva, K. (2018) 'Constructing critical thinking in health professional education', *Perspectives on Medical Education*, 7(3), pp. 156-165.

Kahneman, D. (2011) *Thinking, fast and slow*. 1st ed. edn. New York: New York : Farrar, Straus and Giroux, 2011.

Kahneman, D. and Klein, G. (2009) 'Conditions for Intuitive Expertise: A Failure to Disagree', *The American Psychologist*, 64(6), pp. 515-526.

Kahneman, D., Slovic, P. and Tversky, A. (1982) *Judgment under uncertainty : heuristics and biases*. Edited by Kahneman, D., Slovic, P. and Tversky, A. New York, Cambridge : Cambridge University Press, .

Kahneman, D. and Tversky, A. (1973) 'On the psychology of prediction', *Psychological Review*, 80(4), pp. 237-251.

Kaltura (2020) *Kaltura Capture* (Version 4.3.18) [Computer program]. Kaltura. (Accessed: 8th August 2022)

Kassirer, J. P. (2010) 'Teaching Clinical Reasoning: Case-Based and Coached', *Academic Medicine*, 85(7), pp. 1118-1124.

Kilgour, J. M. and Tayyaba, S. (2015) 'An investigation into the optimal number of distractors in single-best answer exams', *Advances in Health Sciences Education : Theory and Practice*, 21(3), pp. 571-585.

Kitzinger, J. (1995) 'Qualitative Research: Introducing focus groups', *BMJ*, 311(7000), pp. 299-302.

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H. and Wyatt, T. R. (2020) 'Demystifying Content Analysis', *American Journal of Pharmaceutical Education*, 84(1), pp. 127-137.

Koelewijn, G., Hennus, M. P., Kort, H. S. M., Frenkel, J. and van Houwelingen, T. (2024) 'Games to support teaching clinical reasoning in health professions education: a scoping review', *Medical Education Online*, 29(1), pp. 2316971-2316971.

Kogan, J. R., Holmboe, E. S. and Hauer, K. E. (2009) 'Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review', *Journal of the American Medical Association*, 302(12), pp. 1316-1326.

Kononowicz, A. A., Hege, I., Edelbring, S., Sobocan, M., Huwendiek, S. and Durning, S. J. (2020) 'The need for longitudinal clinical reasoning teaching and assessment: Results of an international survey', *Medical Teacher*, 42(4), pp. 457-462.

Kuhn, T. S. (1996) *The structure of scientific revolutions*. 3rd ed. edn. Chicago, London : University of Chicago Press, 1996.

Laing, R. D. (1971) *The politics of the family and other essays*. London: London : Tavistock Publications, 1971.

Lakin, J. M. (2013) 'Sex differences in reasoning abilities: Surprising evidence that male–female ratios in the tails of the quantitative reasoning distribution have increased', *Intelligence* 41(4), pp. 263-274.

Lasater, K. (2007) 'Clinical judgment development: Using simulation to create an assessment rubric', *Journal of Nursing Education*, 46(11), pp. 496-503.

Lee, C.-Y., Lee, C.-H., Lai, H.-Y., Chen, P.-J., Chen, M.-M. and Yau, S.-Y. (2024) 'Bridging theory and practice: a scoping review protocol on gamification's impact in clinical reasoning education', *BMJ open*, 14(12), p. e086262.

Levene, H. (1960) 'Robust tests for equality of variances', *Contributions to Probability and Statistics*, pp. 278-292.

Likert, R. (1932) A technique for the measurement of attitudes. *Archives of Psychology*, 22, pp5-55.

Lineberry, M., Kreiter, C. D. and Bordage, G. (2013) 'Threats to validity in the use and interpretation of script concordance test scores', *Medical Education*, 47(12), pp. 1175-1183.

Lloyd, F. J. and Reyna, V. F. (2001) 'A web exercise in evidence-based medicine using cognitive theory', *Journal of General Internal Medicine*, 16(2), pp. 94-99.

Lopes, D. M., Bregagnollo, G. H., Barbosa, B. M. and Stamm, A. M. N. d. F. (2018) 'The Process of Clinical Reasoning among Medical Students', *Revista Brasileira de Educação Médica*, 42(3), pp. 194-200.

Lumivero (2022) *NVivo* (Version 1.7.1) [Computer program]. QSR International. (Accessed 20th November 2022)

MacLeod, A. (2011) 'Six ways problem-based learning cases can sabotage patient-centered medical education', *Academic Medicine*, 86(7), pp. 818-825.

Malterud, K., Reventlow, S. and Guassora, A. D. (2019) 'Diagnostic knowing in general practice: interpretative action and reflexivity', *Scandinavian Journal of Primary Health Care*, 37(4), pp. 393-401.

Mamede, S., Zandbergen, A., de Carvalho-Filho, M. A., Choi, G., Goeijenbier, M., van Ginkel, J., Zwaan, L., Paas, F. and Schmidt, H. G. (2024) 'Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment', *BMJ Quality and Safety*, pp. bmjqs-2023-016621.

Manetti, W. (2019) 'Sound clinical judgment in nursing: A concept analysis', *Nursing Forum*, 54(1), pp. 102-110.

Marcum, J. A. (2012) 'An integrated model of clinical reasoning: dual-process theory of cognition and metacognition', *Journal of Evaluation in Clinical Practice*, 18(5), pp. 954-961.

Mattingly, C. (1991) 'What is Clinical Reasoning?', *American Journal of Occupational Therapy*, 45(11), pp. 979-986.

McLeod, I., Zhang, Y. and Yu, H. (2003) 'Multiple-Choice Randomization', *Journal of Statistics Education*, 11(1), pp. null-null.

McMillan, S. S., Kelly, F., Sav, A., Kendall, E., King, M. A., Whitty, J. A. and Wheeler, A. J. (2014) 'Using the Nominal Group Technique: how to analyse across multiple groups', *Health Services and Outcomes Research Methodology*, 14(3), pp. 92-108.

McMillan, S. S., King, M. and Tully, M. P. (2016) 'How to use the nominal group and Delphi techniques', *International Journal of Clinical Pharmacy*, 38(3), pp. 655-662.

Mercer, J. (2007) 'The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas', *Oxford Review of Education*, 33(1), pp. 1-17.

Microsoft (2023) *Excel* (Version 2308) [Computer program]. Microsoft Corporation. (accessed 22nd November 2022)

Miles, M. B. and Huberman, A. M. (1984) *Qualitative data analysis : a sourcebook of new methods*. Beverly Hills: Sage Publications.

Miller, G. E. (1990) 'The assessment of clinical skills/competence/performance', *Academic Medicine*, 65(9), pp. S63-S67.

Mjaaland, T. A., Finset, A., Jensen, B. F. and Gulbrandsen, P. (2011) 'Physicians' responses to patients' expressions of negative emotions in hospital consultations: A video-based observational study', *Patient Education and Counselling*, 84(3), pp. 332-337.

Monteiro, S. D., Sherbino, J., Patel, A., Mazzetti, I., Norman, G. R. and Howey, E. (2015) 'Reflecting on Diagnostic Errors: Taking a Second Look is Not Enough', *J Gen Intern Med*, 30(9), pp. 1270-1274.

Morgan, Malani, P. and Diekema, D. J. (2017) 'Diagnostic Stewardship—Leveraging the Laboratory to Improve Antimicrobial Use', *Journal of the American Medical Association*, 318(7), pp. 607-608.

Morgan, D. L. (1988) *Focus groups as qualitative research*. London: London : Sage, c1988.

Morse, J. M. (2006) 'Reconceptualizing Qualitative Evidence', *Qualitative Health Research*, 16(3), pp. 415-422.

MSCAA (2019) *Writing Questions for Undergraduate Exams - A resource pack for MSCAA* [PDF document]. Available at: https://sites.manchester.ac.uk/prime/wp-content/uploads/sites/120/2020/11/MSCAA-Question-Writing-Guide.pdf (Accessed: 5th April 2022).

Murphy, C. M. (2012) 'Writing an Effective Review Article', *Journal of Medical Toxicology*, 8(2), pp. 89-90.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J. and Marteau, T. (1998) 'Consensus development methods, and their use in clinical guideline development', *Health Technology Assessment*, 2(3), pp. i-iv.

Mylopoulos, M., Kulasegaram, K. and Woods, N. N. (2018) 'Developing the experts we need: Fostering adaptive expertise through education', *Journal of Evaluation in Clinical Practice*, 24(3), pp. 674-677.

Neher, J. O., Gordon, K. C., Meyer, B. and Stevens, N. (1992) 'A Five-Step "Microskills" Model Of Clinical Teaching', *The Journal of the American Board of Family Practice*, 5(4), p. 419.

Neistadt, M. E. (1998) 'Teaching clinical reasoning as a thinking frame', *American Journal of Occupational Therapy*, 52(3), pp. 221-229.

Nesbitt, A., Baird, F., Canning, B., Griffin, A. and Sturrock, A. (2013) 'Student perception of workplace-based assessment', *The Clinical Teacher*, 10(6), pp. 399-404.

Neufeld, V. R., Norman, G. R., Feightner, J. W. and Barrows, H. S. (1981) 'Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis', *Medical Education*, 15(5), pp. 315-322.

Neugebauer, R. and Ng, S. (1990) 'Differential recall as a source of bias in epidemiologic research', *Journal of clinical epidemiology*, 43(12), pp. 1337-1341.

Nisbett, R. E., Choi, I., Peng, K. and Norenzayan, A. (2001) 'Culture and systems of thought: Holistic versus analytic cognition', *Psychological Review*, 108(2), pp. 291-310.

Norman, G. (2009) 'Dual processing and diagnostic errors', *Advances in Health Sciences Education*, 14(1 SUPPL), pp. 37-49.

Norman, G., Monteiro, S. and Sherbino, J. (2013) 'Is clinical cognition binary or continuous', *Academic Medicine*, 88(8), pp. 1058-1060.

Norman, G. R., Brooks, L. R., Colle, C. L. and Hatala, R. M. (1999) 'The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning', *Cognition and Instruction*, 17(4), pp. 433-448.

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G. and Mamede, S. (2017) 'The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking', *Academic Medicine*, 92(1), pp. 23-30.

O'Cathain, A. (2010) 'Assessing the Quality of Mixed Methods Research: Toward a Comprehensive Framework' 2. Thousand Oaks: Thousand Oaks: SAGE Publications, Inc, p. 531.

O'sullivan, E. D. and Schofield, S. J. (2018) 'Cognitive bias clinical medicine', *Journal of the Royal College of Physicians of Edinburgh*, 48(3), pp. 225-232.

Ochieng, P. A. (2009) 'An Analysis Of The Strengths And Limitation Of Qualitative And Quantitative Research Paradigms', *Problems of Education in the 21st Century*, 13, p. 13.

Öhlén, J. (2011) 'Review: Janice M. Morse & Linda Niehaus (2009). mixed method design: Principles and procedures', *Forum, Qualitative Social Research*, 12(1).

Onwuegbuzie, A. and Johnson, R. B. (2006) 'The Validity Issue in Mixed Research', *Research in the Schools*, 13(1), p. 48.

Palmer, E. J. and Devitt, P. G. (2007) 'Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper', *BMC Medical Education*, 7(1), pp. 49-49.

Park, W. B., Kang, S. H., Myung, S. J. and Lee, Y.-S. (2015) 'Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students?', *The American Journal of the Medical Sciences*, 350(1), pp. 64-67.

Pelaccia, T., Tardif, J., Triby, E., Ammirati, C., Bertrand, C., Dory, V. M. D. and Charlin, B. M. D. (2014) 'How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews', *Annals of Emergency Medicine*, 64(6), pp. 575-585.

Pelaccia, T., Tardif, J., Triby, E. and Charlin, B. (2011) 'An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory', *Med Educ Online*, 16(1), pp. 5890-9.

Peng, K.-P. and Burish, M. J. (2023) 'Management of cluster headache: Treatments and their mechanisms', *Cephalalgia*, 43(8), doi:10.1177/03331024231196808.

Pennaforte, T., Moussa, A., Loye, N., Charlin, B. and Audetat, M.-C. (2016) 'Exploring a New Simulation Approach to Improve Clinical Reasoning Teaching and Assessment: Randomized Trial Protocol', *JMIR Research Protocols*, 5(1).

Penner, J. C., Schuwirth, L. and Durning, S. J. (2024) 'From Noise to Music: Reframing the Role of Context in Clinical Reasoning', *Journal of General Internal Medicine*.

Pernar, L. I. M., Askari, R. and Breen, E. M. (2020) 'Oral examinations in undergraduate medical education – What is the 'value added' to evaluation?', *The American Journal of Surgery*, 220(2), pp. 328-333.

Poses, P. M. and Isen, A. M. (1998) 'Qualitative research in medicine and health care: Questions and controversy', *Journal of General Internal Medicine*, 13(1), pp. 32-38.

Pottier, P., Hardouin, J. B., Hodges, B. D., Pistorius, M. A., Connault, J., Durant, C., Clairand, R., Sebille, V., Barrier, J. H. and Planchon, B. (2010) 'Exploring how students think: a new method combining think-aloud and concept mapping protocols', *Medical Education*, 44(9), pp. 926-935.

Pradhan, G., Mendinca, N. L. and Kar, M. (2014) 'Evaluation of examination stress and its effect on cognitive function among first year medical students', *Journal of Clinical and Diagnostic Research*, 8(8), pp. BC05-BC07.

Rapezzi, C., Ferrari, R. and Branzi, A. (2005) 'White coats and fingerprints: diagnostic reasoning in medicine and investigative methods of fictional detectives', *British Medical Journal*, 331(7531), pp. 1491-1494.

Rencic, J., Schuwirth, L. W. T., Gruppen, L. D. and Durning, S. J. (2020) 'Clinical reasoning performance assessment: using situated cognition theory as a conceptual framework', *Diagnosis*, 7(3), pp. 241-249.

Resnik, L. and Jensen, G. M. (2003) 'Using clinical outcomes to explore the theory of expert practice in physical therapy', *Physical Therapy*, 83(12), pp. 1090-1106.

Richards, J. B., Hayes, M. M. and Schwartzstein, R. M. (2020) 'Teaching Clinical Reasoning and Critical Thinking: From Cognitive Theory to Practical Application', *Chest*, 158(4), pp. 1617-1628.

Rimmer, A. (2016) 'Ethnic minority and non-UK doctors are more likely to fail exams, GMC data show', *British Medical Journal (Online)*, p. i4100.

Ripp, K. and Braun, L. (2017) 'Race/Ethnicity in Medical Education: An Analysis of a Question Bank for Step 1 of the United States Medical Licensing Examination', *Teaching and Learning in Medicine*, 29(2), pp. 115-122.

Roberts, K., Drennan, V. M. and Watkins, J. (2022) 'Physician associate graduates in England: a cross-sectional survey of work careers', *Future Healthcare Journal*, 9(1), pp. 5-10.

Rowe, G., Wright, G. and Bolger, F. (1991) 'Delphi: A reevaluation of research and theory', *Technological Forecasting & Social Change*, 39(3), pp. 235-251.

Rumsey, D. J. (2003) *Statistics for dummies*. Hoboken, N.J. Great Britain : Wiley, c2003.

Runyan, M. K. (1991) *Reading comprehension performance of learning-disabled and non-learning disabled college and university students under timed and untimed conditions*. ProQuest Dissertations Publishing.

Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J. and 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 and Safety*, 22(8), pp. 672-680.

Salter, I. Y. and Atkins, L. J. (2014) 'What Students Say Versus What They Do Regarding Scientific Inquiry', *Science Education*, 98(1), pp. 1-35.

Sam, A. H., Fung, C. Y., Reed, M., Hughes, E. and Meeran, K. (2022) 'Time for preference-informed foundation allocation?', *Clinical Medicine*, 22(6), pp. 590-593.

Sam, A. H., Hameed, S., Harris, J. and Meeran, K. (2016) 'Validity of very short answer versus single best answer questions for undergraduate assessment', *BMC Medical Education*, 16(1), pp. 266-266.

Sammons, P., Toth, K. and Sylva, K. (2015) *Background to success: differences in A-Level entries by ethnicity, neighbourhood and gender*. London: University of Oxford/Sutton Trust.

Saposnik, G., Redelmeier, D., Ruff, C. C. and Tobler, P. N. (2016) 'Cognitive biases associated with medical decisions: a systematic review', *BMC Medical Informatics and Decision Making*, 16(1), pp. 1-14.

Sardesai, N., Russo, P., Martin, J. and Sardesai, A. (2024) 'Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training', *Postgraduate Medical Journal*.

Schmidt, H. G., Norman, G. R. and Boshuizen, H. P. A. (1990) 'A cognitive perspective on medical expertise: Theory and implications', *Academic Medicine*, 65(10), pp. 611-621.

Schuwirth (2009) 'Is assessment of clinical reasoning still the Holy Grail?', *Medical Education*, 43(4), pp. 298-300.

Schuwirth, L., van der Vleuten, C. and Durning, S. J. (2017) 'What programmatic assessment in medical education can learn from healthcare', *Perspectives on Medical Education*, 6(4), pp. 211-215.

Shabanowitz, N., Nelson, N. R., Rodgers, J. E. and Rhoney, D. H. (2024) 'Student Pharmacists Provide Similar Quality Clinical Reasoning Feedback as Resident Teaching Assistants', *American Journal of Pharmaceutical Education*, 88(4), pp. 100677-100677.

Shenwai, M. R. and Patil, K. B. (2013) 'Introduction of structured oral examination as a novel assessment tool to first year medical students in physiology', *Journal of Clinical and Diagnostic Research*, 7(11), pp. 2544-2547.

Shibley Hyde, J. and Linn, M. C. (2006) 'Diversity - Gender similarities in mathematics and science', *Science*, 314(5799), pp. 599-600.

Shin, S. J., Kim, K. S. and Lee, D. S. (2011) 'The effect of personal character on the results of clinical performance skill tests', *Korean Journal of Medical Education*, 23(2), pp. 111-117.

Siegelman, J., Bernstein, L., Goedken, J., Lewin, L., Schneider, J., Ward, M. and Stoddard, H. (2024) 'Assessment of Clinical Reasoning During a High Stakes Medical Student OSCE', *Perspectives on Medical Education*, 13(1), p. 629.

Simmons, B. (2010) 'Clinical reasoning: concept analysis', *Journal of Advanced Nursing*, 66(5), pp. 1151-1158.

Simon, H. A. (1955) 'A Behavioral Model of Rational Choice', *The Quarterly Journal of Economics*, 69(1), pp. 99-118.

Skellern, C. (2020) 'Thinking fast and slow in the evaluation of injury plausibility in child protection', *Journal of Paediatrics and Child Health*, 56(9), pp. 1330-1334.

Slonim-Nevo, V. and Nevo, I. (2009) 'Conflicting Findings in Mixed Methods Research: An Illustration From an Israeli Study on Immigration', *Journal of Mixed Methods Research*, 3(2), pp. 109-128.

Smith, M. M., Sherry, S. B., Chen, S., Saklofske, D. H., Mushquash, C., Flett, G. L. and Hewitt, P. L. (2018) 'The perniciousness of perfectionism: A meta-analytic review of the perfectionism–suicide relationship', *Journal of Personality*, 86(3), pp. 522-542.

Sockalingam, S. M. D. M. H. P. E., Mulsant, B. H. M. D. M. S. and Mylopoulos, M. P. D. (2016) 'Beyond Integrated Care Competencies: The Imperative for Adaptive Expertise', *General Hospital Psychiatry*, 43, pp. 30-31.

Speicher, T. E., Bell, A., Kehrhahn, M. and Casa, D. J. (2012) 'Case-Based Analogical Reasoning: A Pedagogical Tool for Promotion of Clinical Reasoning', *Athletic Training Education Journal*, 7(3), pp. 129-136.

Steginga, S. K. and Occhipinti, S. (2004) 'The Application of the Heuristic-Systematic Processing Model to Treatment Decision Making about Prostate Cancer', *Medical Decision Making*, 24(6), pp. 573-583.

Stempsey, W. E. (2009) 'Clinical reasoning: New challenges', *Theoretical Medicine and Bioethics*, 30(3), pp. 173-179.

Stewart, J. (2001) 'Is the Delphi technique a qualitative method?', *Medical Education*, 35(10), pp. 922-923.

Stolper, E., Van Bokhoven, M., Houben, P., Van Royen, P., Van De Wiel, M., Van Der Weijden, T. and Jan Dinant, G. (2009) 'The diagnostic role of gut feelings in general practice A focus group study of the concept and its determinants', *BMC Family Practice*, 10(1), pp. 17-17.

Sullivan, G. M. and Artino, J. A. R. (2013) 'Analyzing and interpreting data from likert-type scales', *Journal of Graduate Medical Education*, 5(4), pp. 541-542.

Tashakkori, A. and Teddlie, C. (2010) *Sage handbook of mixed methods in social & behavioral research*. Second edition. edn. Los Angeles: Los Angeles : SAGE, 2010.

Tavakol, M. and Sandars, J. (2014a) 'Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I', *Medical Teacher*, 36(9), pp. 746-756.

Tavakol, M. and Sandars, J. (2014b) 'Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part II', *Medical Teacher*, 36(10), pp. 838-848.

Teddlie, C. (2009) *Foundations of mixed methods research : integrating quantitative and qualitative approaches in the social and behavioral sciences*. Edited by Tashakkori, A. Los Angeles: Los Angeles : SAGE, c2009.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L. and Goldsmith, C. H. (2010) 'A tutorial on pilot studies: The what, why and how', *BMC Medical Research Methodology*, 10(1), pp. 1-1.

Thammasitboon, S., Rencic, J. J., Trowbridge, R. L., Olson, A. P. J., Sur, M. and Dhaliwal, G. (2018) 'The Assessment of Reasoning Tool (ART): structuring the conversation between teachers and learners', *Diagnosis*, 5(4), pp. 197-203.

Thampy, H., Willert, E. and Ramani, S. (2019) 'Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid', *Journal of General Internal Medicine*, 34(8), pp. 1631-1636.

Thomson, O. P., Petty, N. J. and Moore, A. P. (2014) 'Clinical decision-making and therapeutic approaches in osteopathy - A qualitative grounded theory study', *Manual Therapy*, 19(1), pp. 44-51.

Tolsma, R., Shebrain, S., Berry, S. D. and Miller, L. (2024) 'Medical student perceptions of assessments of clinical reasoning in a general surgery clerkship', *BMC Medical Education*, 24(1), pp. 211-211.

Tonelli, M. R. and Shapiro, D. (2020) 'Experiential knowledge in clinical medicine: use and justification', *Theoretical medicine and bioethics*, 41(2-3), pp. 67-82.

Tsiga, E., Panagopoulou, E., Sevdalis, N., Montgomery, A. and Benos, A. (2013) 'The influence of time pressure on adherence to guidelines in primary care: an experimental study', *BMJ Open*, 3(4), p. e002700.

Tversky, A., Tversky, A., Kahneman, D. and Kahneman, D. (1974) 'Judgment under uncertainty: heuristics and biases. Biases in judgments reveal some heuristics of thinking under uncertainty', *Science*, 185(4157), pp. 1124-1131.

UCAS (2022) *UCAS Undergraduate sector-level end of cycle data resources 2022*. Available at: https://www.ucas.com/data-and-analysis/undergraduate-statistics-and-reports/ucas-undergraduate-end-cycle-data-resources-2022 (Accessed: 9th August 2024).

Unsworth, C. A. (2004) 'Clinical reasoning: How do pragmatic reasoning, worldview and client-centredness fit?', *British Journal of Occupational Therapy*, 67(1), pp. 10-19.

Upshur, R. (1997) 'Certainty, probability and abduction: why we should look to C.S. Peirce rather than Godel for a theory of clinical reasoning', *Journal of Evaluation in Clinical Practice*, 3(3), pp. 201-6.

Van de Ven, A. H. and Delbecq, A. L. (1972) 'The nominal group as a research instrument for exploratory health studies', *American Journal of Public Health*, 62(3), pp. 337-342.

Van der Vleuten, C., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J. and van Tartwijk, J. (2012) 'A model for programmatic assessment fit for purpose', *Medical Teacher*, 34(3), pp. 205-214.

Van der Vleuten, C. P. M. and Schuwirth, L. W. T. (2019) 'Assessment in the context of problem-based learning', *Advances in Health Sciences Education : Theory and Practice*, 24(5), pp. 903-914.

van Teijlingen, E. and Hundley, V. (2002) 'The importance of pilot studies', *Nursing Standard*, 16(40), pp. 33-36.

Van Teijlingen, E. R., Rennie, A.-M., Hundley, V. and Graham, W. (2001) 'The importance of conducting and reporting pilot studies: the example of the Scottish Births Survey', *Journal of Advanced Nursing*, 34(3), pp. 289-295.

Veen, M. (2021) 'Creative leaps in theory: the might of abduction', *Advances in Health Sciences Education : Theory and Practice*, 26(3), pp. 1173-1183.

Vertue, F. M. and Haig, B. D. (2008) 'An abductive perspective on clinical reasoning and case formulation', *Journal of Clinical Psychology*, 64(9), pp. 1046-1068.

Vrijsen, B. E. L., Naaktgeboren, C. A., Vos, L. M., van Solinge, W. W., Kaasjager, H. A. H. and ten Berg, M. J. (2020) 'Inappropriate laboratory testing in internal medicine inpatients: Prevalence, causes and interventions', *Annals of Medicine and Surgery*, 51, pp. 48-53.

Walsh, J. L., Harris, B. H. L. and Smith, P. E. (2017) 'Single best answer question-writing tips for clinicians', *Postgraduate Medical Journal*, 93(1096), pp. 76-81.

Ward, T., Clack, S. and Haig, B. D. (2016) 'The Abductive Theory of Method: Scientific Inquiry and Clinical Practice', *Behaviour Change*, 33(4), pp. 212-231.

Ward, T. and Haig, B. (1997) 'Abductive reasoning and clinical assessment', *Australian Psychologist*, 32(2), pp. 93-100.

Weaver, S. J., Newman-Toker, D. E. and Rosen, M. A. (2012) 'Reducing Cognitive Skill Decay and Diagnostic Error: Theory-Based Practices for Continuing Education in Health Care', *Journal of Continuing Education in the Health Professions*, 32(4), pp. 269-278.

Wellbery, C. (2010) 'The art of medicine The value of medical uncertainty?', *The Lancet (British edition)*, 375(9727), pp. 1686-1687.

Whitbeck, C. (1981) 'What is diagnosis? Some critical reflections', *Metamedicine*, 2(3), pp. 319-329.

Williams, G. and Nottingham, S. (2022) 'Using the Unfolding Case Study to Improve Clinical Reasoning', *The Internet Journal of Allied Health Sciences and Practice*.

Young, J. Q., Van Merrienboer, J., Durning, S. and Ten Cate, O. (2014) 'Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86', *Medical Teacher*, 36(5), pp. 371-384.

Young, M., Thomas, A., Lubarsky, S., Ballard, T., Gordon, D., Gruppen, L. D., Holmboe, E., Ratcliffe, T., Rencic, J., Schuwirth, L. and Durning, S. J. (2018) 'Drawing boundaries: The difficulty in defining clinical reasoning', *Academic Medicine*, 93(7), pp. 990-995.

Zhang, Q., Wang, C., Zhao, Q., Yang, L., Buschkuehl, M. and Jaeggi, S. M. (2019) 'The malleability of executive function in early childhood: Effects of schooling and targeted training', *Developmental Science*, 22(2), pp. e12748-n/a.

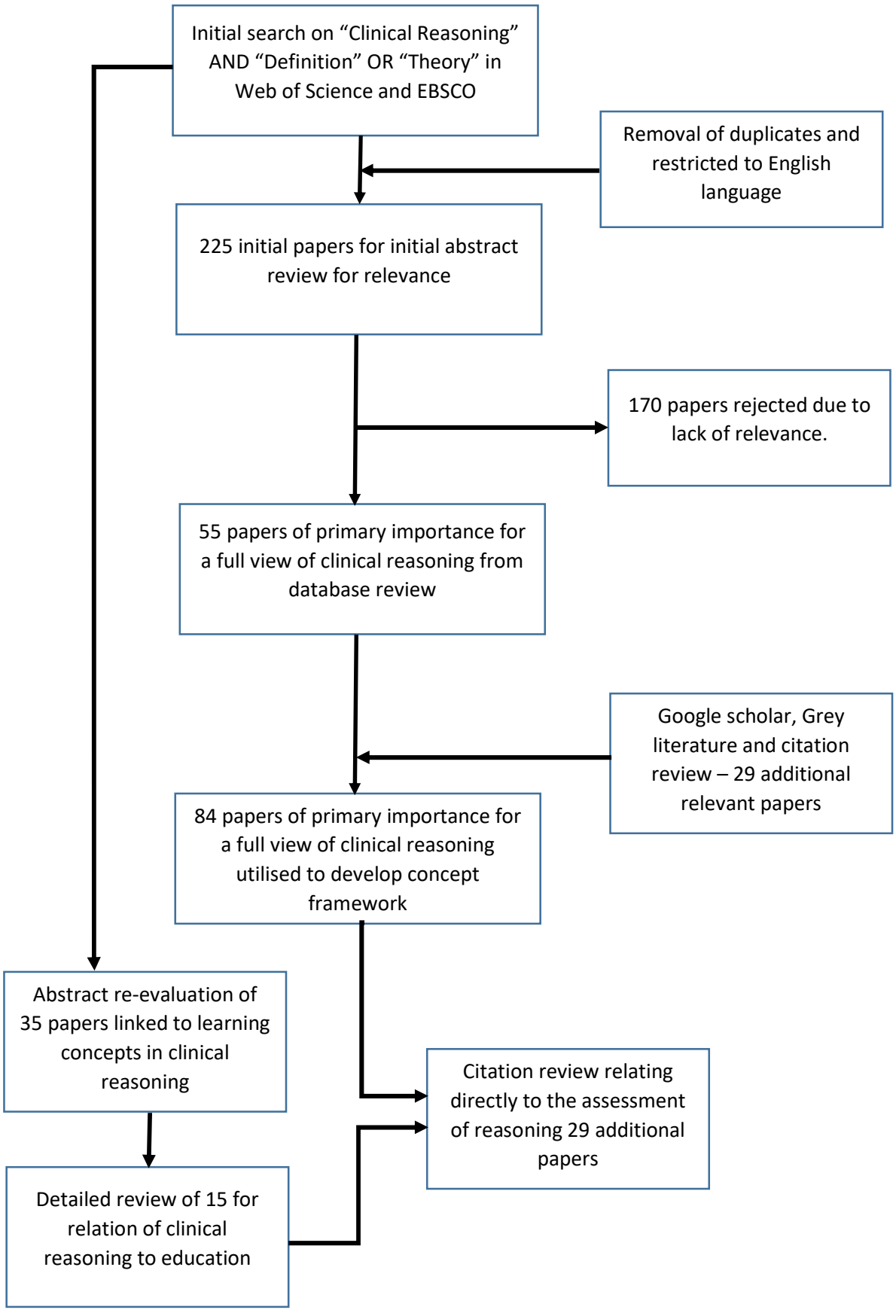
Zheng, J., Li, S. and Lajoie, S. P. (2022) 'Diagnosing virtual patients in a technology-rich learning environment: a sequential Mining of Students' efficiency and behavioral patterns', *Education and information technologies*, 27(3), pp. 4259-4275.

Outputs

Study and provisional results poster presented at the Sheffield school of medicine PGR research day (4th July 2023)

Full study presented at the 2024 Physician Associate Educators Conference (23rd May 2024)

Appendix 1 – Narrative Review Search Strategy



Appendix 2 – Detailed paper review, example

Authors	Title	Sample	Phenomenon of Interest	Design	Evaluation	Research Type	Notes
Artino, A. R., et al.	Exploring clinical reasoning in novices: a self-regulated learning microanalytic assessment approach	71 second year medical students	Regulatory processes of medical students	SRL microanalytic assessment	Descriptive statistics and hierarchical linear regression	Mixed Methods	Deficits in self-regulated learning are predictors of a range of performance indicators. SRL is teachable. SRL methods assume that student's thoughts, feelings and actions are context bound and fluctuate across tasks. SRL microanalysis s structured interviewing aimed to capture students thoughts, feelings and actions as they participate in different tasks. Did not look at the skill but rather the individual's regulation. This may be a very useful technique for my research interest
Berger, A. J. B. A., et al.	Assessment of medical student clinical reasoning by "lay" vs physician raters: inter-rater reliability using a scoring guide in a multidisciplinary objective structured clinical examination.	25 clinical notes	Whether lay assessors can assess clinical reasoning from clinical notes	Case-control study	Inter-rater reliability	Qualitative	Note writing to be common in many OSCEs as a structured patient note then evaluated by expert physicians. Use rubric construction to support lay-rater decision-making. Notable that alpha was low for both physician and lay raters thus domains of reasoning appear relatively independent. But this does not evaluate how students reason.
Bowen, J. L. (2006).	Educational Strategies to Promote Clinical Diagnostic Reasoning.	N/A	Clinical reasoning process and implications for education	Literature review (unstructured)	Application of theory to fictional case study	N/A	Mention the idea of asking learners to articulate their reasoning processes. Comments on the transformation of patient specific details into abstract terms. Suggests that the expert compares and contrasts multiple scripts. Suggests novices use more analytical. Promote think aloud and justification of decision processes. Need to encourage "compare and contrast" reading looking at two hypotheses at a time.
Crossley, J., et al.	Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales	>2000 med trainees and >4000 assessors with 24322 assessments	WBA methods and alternative rating scales	Multicentre trial of medical trainees	Statistical evaluation of judgement scores	Quantitative	Note that good clinical practice is dependent upon good assessment practice. Discuss miniCEX, CBD and ACAT. Better alignment of the judgement question improved the number of assessments needed to reach an effective judgement. Constructive alignment of rating scale with the core judgement needed improves assessor decision-making.

Daniel, M., et al.	Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance.	N/A	Assessment of Clinical Reasoning	Structured lit review	Constructivist review	N/A	This is a practical guide to assessment methods as a reference for educators noting it is not exhaustive. Only looked at those relating to medics. Three areas are non-WBA, simulated clinical environments and WBAs. Considers MCQ/SBA and KFEs to not be generally representative of CR in practice thus cannot be relied upon to ensure successful skills transfer. Simulated (e.g. OSCE) have reasonable alignment to practice but context is controlled and performance correlation to non-WBA methods is generally poor. Also resource and SP/Rater dependent. WBAs rely on real patients as stimuli. reasonable content and response process validity. Ideally need multiple raters across multiple times and cases. Need to ensure an appropriate range of assessments and bear in mind that non-WBA assessments permit broad sampling.
Durning, S. J., et al.	The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning.	End-of-second-year medical students	Feasibility, reliability and validity evidence of a post encounter form for assessing clinical reasoning.	PEF form completion within 3-part OSCE stations with correlation analysis	Correlation and feasibility analyses	Qualitative	Seeks to explicitly evaluate the use of a PEF to assess CR post OSCE in pre-clinical students. Showed good inter-rater reliability. Potential to apply through "2 station" OSCE approach in our current system where the second part is a written evaluation PEF to explore reasoning. This case was diagnosis specific. Could also add a question as to describe how you reached this diagnosis which might evaluate the thinking process. Note the potential to explore for trainees at other levels.
Evans, B. J., et al.	Effects of communication skills training on students' diagnostic efficiency	60 medical students	Effect of communication skills training on diagnostic ability	Review of video recorded consultations case-control	Observer ratings for the two groups	Qualitative	Note criticisms that much of student time focusses on diagnostic skill to the detriment of therapeutics. Demonstrated that improved communication skills led to the acquisition of more diagnostically relevant information. The improved outcomes were not at the sacrifice of time - there was no increase in the time required.
Gingerich, A., et al.	Rater-based assessments as social judgments: rethinking the aetiology of rater errors.	N/A	Social judgement formation linked to inter-rater reliability	Literature review (structured)	Discussion of the literature	Literature review	Note evidence of clear differences through the concept of person models. Deliberate attempts to alter categorisation-based assumptions may have the opposite effect. Issue of nominal vs Ordinal measurement (arguably the "purest" definition of the use of a global scale). Suggests that we need to tease apart error unintentionally attributed to human biases and error. Links to Crossley et al work. How does nominal vs ordinal judgements align?

Gruppen, L. D., et al.	Enhanced Requirements for Assessment in a Competency-Based, Time-Variable Medical Education System.	N/A	Competency based medical education	Literature review (unstructured)	Discussion and opinion	N/A	Time variability for competence acquisition puts particular demand on the assessment process. Commenting on competency aligns with Crossley et al and changing the rating scales that we utilise. CBME cannot work with purely scheduled examinations. To undertake requires significantly more management support and data sharing. Challenge of context specificity. Multiple formative assessments may enable us to make summative judgements. Entrustment decisions (e.g. clinical skills) add an element to this process
Hodges, B. and J. H. McIlroy	Analytic global OSCE ratings are sensitive to level of training	19 yr 3 and 38 yr4 clinical clerks	Construct validity of analytic global rating scales	Case-control	Comparison of checklist and global score methods	Quantitative	Note evidence that global ratings appear to have psychometric properties at least as good as, and often better than, those of checklists. Checklists may also discriminate against those who are better at gathering data. One examiner assessed students on two different forms. 4 global ratings then summated to an overall global rating. Found global ratings to have higher reliability and construct validity.
Hrynchak, P., et al.	Key-feature questions for assessment of clinical reasoning: a literature review	N/A	Key-feature questions reliability and validity	Literature review	Discussion and opinion	Literature review	Results of Key feature question exams have been shown to predict future physician performance. Note SCT and CRP concerns about reliability in research and time-heavy to develop but can produce reliable exams. KFQ concept is that only a few key elements are needed in the resolution of any clinical problem. Approach to testing rather than format defined. Usually brief case description with KF embedded followed by 1 or more questions (usually 2 or 3) targeted at decision-making rather than factual knowledge. 40 cases needed to produce a 0.8 reliability - reliability improved by using 2-3 items per case. Suggestion that they assessment itself did not test reasoning. Use of lay language improves discrimination between candidates. SBAs are KFQs although in a limited way.
Ilgen, J. S., et al.	Assessing Diagnostic Reasoning: A Consensus Statement Summarizing Theory, Practice, and Future Needs	N/A	Assessment of reasoning skills in Emergency Physicians	Narrative from AEM consensus conference	N/A	N/A	Little new in here - key point is the final statement that multiple strategies must be used if an accurate assessment is to be gained as no single strategy is valid alone. Programme level assessment is therefore crucial

John, J. N., et al.	The Mini-CEX: A Method for Assessing Clinical Skills	1228 min-CEX encounters	To evaluate the Mini-CEX	Review of mini-CEX assessments	Multi-level analysis	Quantitative	Conclude that Mini-CEX through multiple encounters evaluates across multiple problems and settings. Also evaluated by quarter to assess for development change. Complexity had a small but significant correlation with examiner ratings. Felt to be superior to the traditional CEX - no specific comment on its contribution to reasoning.
Kogan, J. R., et al.	Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review.	85 studies	Tools for direct observation of clinical skills	Literature review	Systematic review	Systematic review	Observers were infrequently trained to use the various tools and comments that rater training is generally unknown. Faculty development is important. Development of expertise requires accurate and detailed assessment and feedback. Mini-CEX seen as highly valid and reliable. Notes only a few methods reviewed demonstrated sufficient validity to warrant extensive use.
Lee, A., et al.	Using illness scripts to teach clinical reasoning skills to medical students	53 fourth year medical students	Whether an intervention with illness scripts improved reasoning	Case-control (24-29)	Comparison of DTI and CRP scores between groups	Mixed methods	Intervention group had tailored teaching including use of NEJM case studies. It does feel that the DTI is being used in a way it was never intended for here and showed no difference. Clear change was shown in CRP exercise - the 8w gap to this being performed adds validity in reducing mastery illusion effects. Suggests teaching through illness scripts linked to case studies has potential to improve reasoning.
Lineberry, M., et al.	Threats to validity in the use and interpretation of script concordance test scores	SCT reports in the literature	Validity of SCT tests	Literature review	Literature review	Literature review	Were able to demonstrate improvement in scores simply by not selecting extreme end options (or even just the middle option) and note the omission of certain measurement errors. Suggest alteration to mark schemes noting that no SCT reports use empirical data to justify their scoring keys.
Neher, J. O., et al.	A Five-Step "Microskills" Model Of Clinical Teaching.	(response to experience 23 respondents)	Microskills in teaching behaviours	Feedback on model	feedback from educators	Evaluation of method	Note that teaching interaction in time pressured circumstances must be necessarily brief. 5 microskills - get commitment, probe for supporting evidence, teach general rules, reinforce what was done right and correct mistakes. Clear potential to assess reasoning in the second stage.
Nendaz, M. R., et al.	Teaching Diagnostic Skills: Clinical Vignettes or Chief Complaints?	42 medical students, 53 residents and 60 gen internists	Vignettes with all diagnostic info vs chief complaint only	Case-control	Comparison of diagnostic information between groups	Mixed Methods	Use of vignettes may not enable learners to integrate data acquisition skills into their learning. 2 problems selected and presented either as a vignette or chief complaint. Correct diagnosis was higher with vignettes - in chief compliant those getting it correct obtained more data than was contained in the vignette. Underlines the critical nature of communication and information

							gathering in reasoning "how to obtain and organise the information should not be trained in dissociation from its diagnostic application". Limited by only 2 cases studied
Palmer, E. J. and P. G. Devitt	Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple-choice questions?	50MCQs and 139 MEQ stages	Effectiveness of MEQs in testing higher order functions	Question review	Against blooms taxonomy	Qualitative	Evidence that MEQs can test higher order functions if well written. Used a modified blooms taxonomy to judge the questions. Used good exam writing options to aid. In this evaluation MCQs were better at testing higher level skills. Many MEQs simply measured recall and fact listing. Study suggests MCQs alone can test broad curriculum and cognitive skills. May be linked to poor writing skills - criticism of MCQs may be more related to poor construction.
Park, W. B., et al.	Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students?	65 fourth year medical students	Whether OSCE scores reflect student reasoning ability	Student written response evaluation	Scoring of responses compared to wider educational metrics	Quantitative	Between stations students undertook a 5-minute written exercise of DDs and the information that supported them this was then assessed. Reasoning scores were calculated by the number of diagnoses and the components and then totalled. Reasoning score was not correlated with OSCE score or knowledge test score but was correlated with GPA and diagnostic accuracy score. Suggests OSCE could not differentiate students reasoning skills. Checklist marking is very limited in its reflection of reasoning ability.
Pottier, P., et al.	Exploring how students think: a new method combining think-aloud and concept mapping protocols	3/4/5 yr med students - numbers unclarified	Reasoning processes of students through talk aloud processes	Quasi-experimental combining a new think-aloud protocol and new scoring method	Reliability and validity assessments	Quantitative	Note challenge in assessment - can a reasoning process be considered "bad" if the problem is solved. These were theoretical problems not real patient cases thus removing context from the equation. Very complex and requires significant training to undertake successfully. Note that pattern recognition may not be captured in this process. No real comments on the time to perform
Raupach, T., et al.	Test-enhanced learning of clinical reasoning: a crossover randomised trial	125 med students (87 complete data)	repeat testing with key feature questions v Case based learning	RCT	Scores in evaluations	Quantitative	Combined computer-based learning with key-feature questions. Identical groups but one had test questions - all formative. Particular interest in the retention tests. I am unclear how this is specific to testing reasoning which is suggested yet they acknowledge the focus as being knowledge retention. Notes e-learning as an alternative to SPs but does loss of context make findings relatively invalid?

Schipper, S. and S. Ross	Structured teaching and assessment A new chart-stimulated recall worksheet for family medicine residents	N/A	Chart stimulated recall	N/A	N/A	N/A	Chart-simulated recall is good for "early-closers". Suggest can aid reasoning development. No actual evaluation or attempt to evaluate
Schuwirth, L.	Is assessment of clinical reasoning still the Holy Grail?	N/A	Assessment of reasoning	N/A	N/A	Opinion	Notes in many simulation-based tests intermediates outscore experts - raises validity concerns. Illness script theory demonstrates that a test that rewards thoroughness and penalises efficiency is invalid. Note that is capturing assessment of process that remains the challenge not outcome. We should not seek to disentangle reasoning from knowledge. Suggests the need for triangulation across multiple assessment methodologies

Appendix 3 – Ethics Approval for Nominal Group Process

Downloaded: 14/05/2022
Approved: 28/10/2021

James Gray
Registration number: 200278917
Medical School
Programme: FCM MDH Postgraduate Research Programme

Dear James

PROJECT TITLE: Understanding how healthcare students undertake clinical reasoning in examinations - Part 1, Nominal Group Technique to determine questions for the mixed methods part of the study
APPLICATION: Reference Number 043363

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 28/10/2021 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 043363 (form submission date: 27/10/2021); (expected project end date: 01/11/2021).
- Participant information sheet 1098480 version 1 (27/10/2021).
- Participant information sheet 1097226 version 1 (01/09/2021).
- Participant information sheet 1097227 version 2 (09/09/2021).
- Participant consent form 1097228 version 1 (01/09/2021).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Charlotte Cole
Ethics Administrator
Medical School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Appendix 4 – Information for participants in the Nominal Group Process

1. Research Project Title:

Understanding how students undertake clinical reasoning in assessments – a mixed methods approach.

This part is an initial Nominal Group technique approach to define the key questions for the subsequent mixed methods research into this topic as part of a PhD

2. Invitation paragraph

You are being invited to take part in the initial part of a research project. Before you decide 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 me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.'

3. What is the project's purpose?

Clinical reasoning has been recognised as a key skill required for clinicians for which greater emphasis should be provided within clinical curricula. Review of the literature shows little to consider how students undertake reasoning within examinations, specifically OSCE and SBA examinations despite the fact that a better understanding of this can feed into the way that reasoning is taught within clinical curricula.

This initial part of the project is designed to scope the questions that will be used for both the quantitative and qualitative parts of the study.

4. Why have I been chosen?

Nominal group technique relies on the use of appropriate experts in the field in order to gain the best possible outcome. You have been identified as having an interest in the topic and this research project and have been chosen for your knowledge on clinical reasoning within the Medical educations setting.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting you in any way. You do not have to give a reason.

6. What will happen if I choose to take part?

If you choose to take part, you will be provided with some background reading prior to the agreed date of the group and detailed guidance as to the process to be undertaken. You will be expected to attend to be part of the group which should take around three hours and will be run online. This will be a single, one off, commitment.

7. What do I have to do?

You are required to read the background information to help understand the context of the outputs that are being sought. You will then be asked to participate as an active member of the group following a pre-defined structure as set out in the accompanying document.

8. What are the possible disadvantages and risks of taking part?

There are no foreseeable disadvantages or risks of taking part.

9. What are the possible benefits of taking part?

The aim of this part of the project as a whole is to help us to define key questions that can help us better understand the process of reasoning that students undertake in written and OSCE examinations.

10. What if something goes wrong?

Should you have any complaints about the process this should be addressed to Dr James Gray the lead researcher. Should any response not be satisfactory concerns can be raised to Professor Michelle Marshall as the PhD supervisor for Dr Gray.

11. Will my taking part in this project be kept confidential?

Given the nature of the project we would seek to acknowledge your part in the NGT process and include your details within acknowledgements of any subsequent presentations or publications. Should you wish to remain completely confidential this will be assured with all identifiable details removed.

12. What will happen to the results of the research project?

The results of the NGT will feed into the qualitative (focus groups) and quantitative (Likert scale) arms of the mixed methods research project. Depending on the outputs we may seek to publish the results of the NGT as a short report.

13. Who is organising and funding the research?

This project is supported by the Academic Unit of Medical Education within the University 1 and has been ethically approved via the University's Research Ethics Committee who monitor the application and delivery of the University's Ethics Review Procedure across the University.

14. Will I be recorded, and how will the recorded media be used?

The process will be recorded, and the recording kept securely for use purely as reference for the researcher and supervisors. There will be no external use of the recording.

15. Contact for further information

Dr James Gray – Course Director Physician Associate Studies

Email: mailto:j.t.gray@sheffield.ac.uk

07588 659106

Appendix 5 – Detail of process for Nominal Group Process participants

1. Pre-group requirements

Prior to the day of the group you will be provided with some pre-reading. This will be a detailed review of the topic and conceptualisation of reasoning that forms the background to the work.

Secondly you should start to think about the process and how you might formulate statements that will be used as questions in both the quantitative and qualitative arms of the study. As detailed below.

2. Completion of a consent form online

In order to ensure contemporaneous signing of consent from you a google form will be set up to capture consent on the day – this ensures consent is witnessed, contemporary and fully informed.

3. Stage 1 - Understanding the way the information generated is to be used

The study has a quantitative and qualitative arm. In the quantitative arm students will be asked to sit a series of SBA questions then immediately answer questions on their reasoning process or sit OSCE stations then immediately be asked the same. It is expected that these will a series of easy to complete Likert scales. In the qualitative arms the statements will form the basis of questions for focus groups of students.

The purpose of the NGT is to use your expertise to help determine the most important questions to ask to understand reasoning. These should be formulated in terms such as "The context setting of the question has a large impact on my decision-making". Such a statement can then be used for a Likert scale but also made into a focus group question i.e. "Does the context setting of a question have a large impact on your decision-making"

The purpose of the NGT is to formulate and then rank the questions to ensure that those which are most important are asked. The initial stage is therefore an introduction and discussion to ensure that all participants are aware of the process and its aims.

4. Stage 2 – Silent Idea Generation

The second stage of the process is silent idea generation. Each participant will write down the questions that they feel most appropriate to answer the question. With this being online it is advised that these are written on a word document to screen share.

5. Stage 3 – Round Robin listing of items

In stage 3 we will go round the group reading out one question at a time. No rationale is required at this point, and this is repeated until all members have presented all of their questions. The researcher will collate the questions on a single document for screen sharing in stage 4.

6. Stage 4 – Group discussion and clarification

Stage 4 involves a brief discussion for clarification of the initial items generated to ensure that all group members have a common understanding of the questions presented. At this point any duplicates (as agreed in discussion) will be removed.

7. Stage 5 – Ranking and Ordering items

In this stage you are asked to individually rank the top 10 questions from 1 (most important) to 10 (least important) that require asking in order to gain a better understanding of student reasoning during assessment. This is again done "silently".

The ranks for each question are then summed along with the number voting for each question (to aid discussion) and recorded in a spreadsheet. Discussion can then be had around any items which are closely correlated and, if considered appropriate a re-ranking process undertaken.

8. Stage 6 – Presentation of the final list and discussion

The final list will then be confirmed and presented, and discussion had regarding comfort with the outcome. One discussion will be whether ten questions is too many, especially if lower ranked questions are thought to add limited value.

9. Post Group Discussion

At the end of the discussion your role as a participant will be at an end however we will keep you up to date with the project as a whole. Should the work be publishable as a short report you would receive an acknowledgement in any publication should you consent to this. You are able to remain anonymous following the process should you wish. Your personally identifiable details will only be kept on secure storage separate from the outcome data until the project ends in 2025. It will then be destroyed.

10. What will be needed to participate

You will need a computer with the ability to undertake a video meeting. The meeting will be recorded and only reviewed by the researcher and their supervisor.

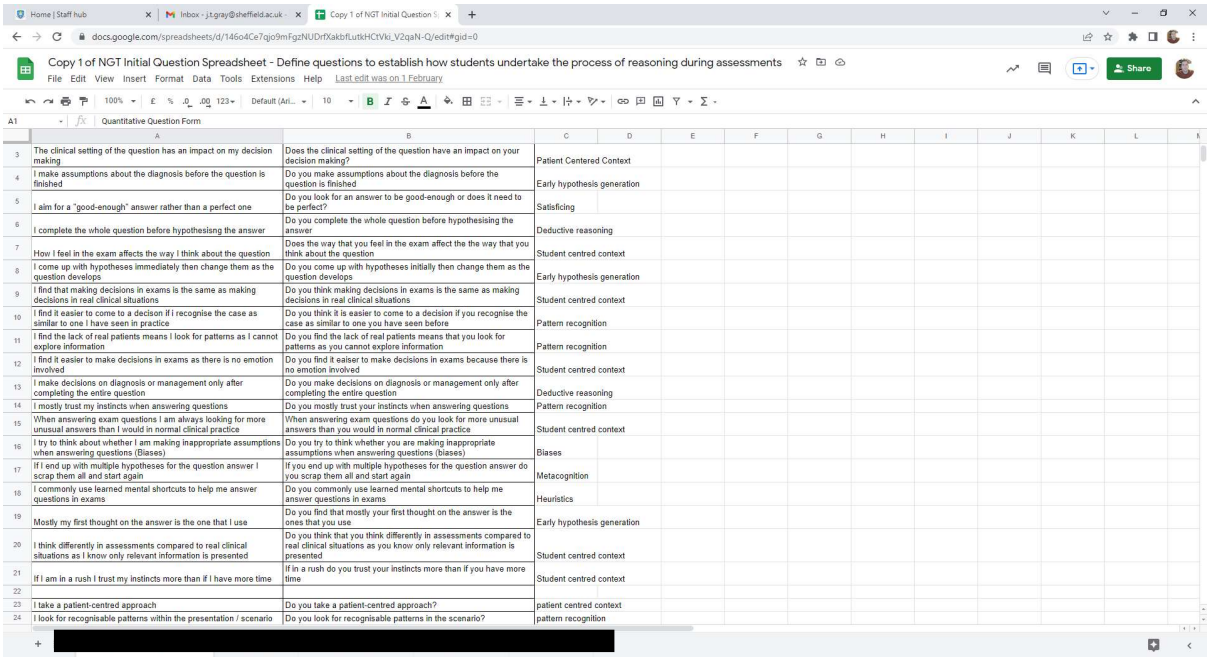
11. Contact for further information

Dr James Gray – Senior University teacher

Email: mailto:j.t.gray@sheffield.ac.uk

07588 659106

Appendix 6 – Screenshots from the Nominal Group Process



Copy 1 of NGT Initial Question Spreadsheet - Define questions to establish how students undertake the process of reasoning during assessments

File Edit View Insert Format Data Tools Extensions Help Last edit was on 1 February

Share

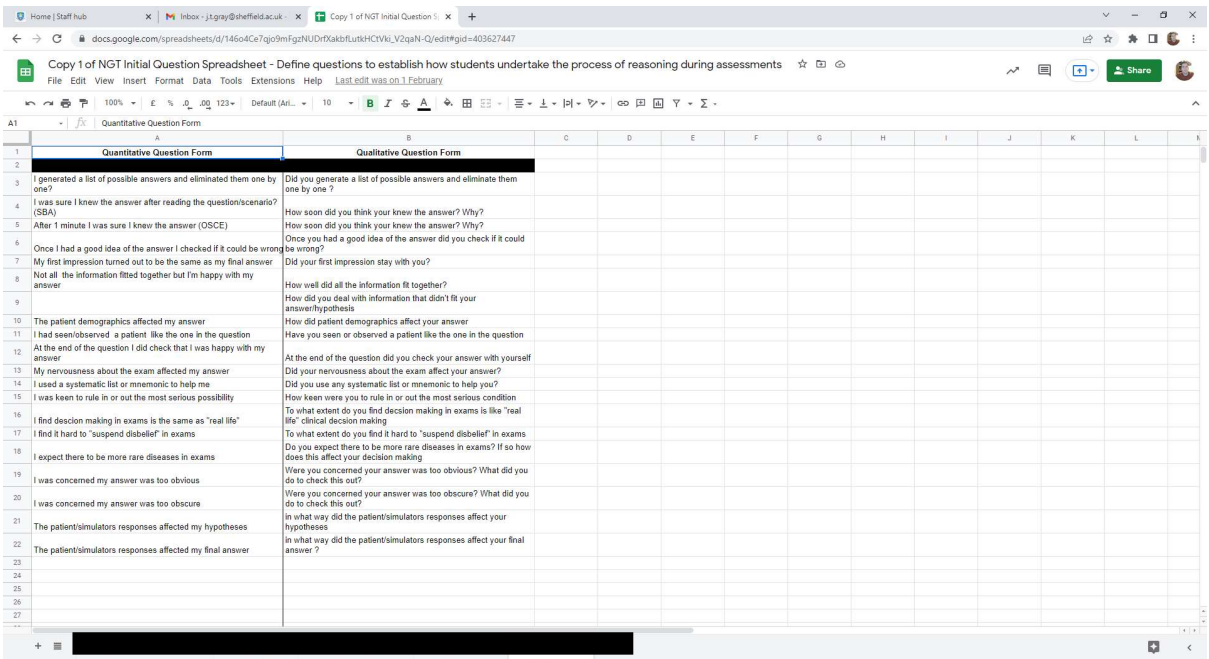
docs.google.com/spreadsheets/d/146o4Ce7qjo9mFgzNUDrfXakbfLutkHCtVki_V2qaN-Q/edit#gid=0

A1 Quantitative Question Form

	A	B	C
3	The clinical setting of the question has an impact on my decision making	Does the clinical setting of the question have an impact on your decision making?	Patient Centered Context
4	I make assumptions about the diagnosis before the question is finished	Do you make assumptions about the diagnosis before the question is finished	Early hypothesis generation
5	I aim for a "good-enough" answer rather than a perfect one	Do you look for an answer to be good-enough or does it need to be perfect?	Satisficing
6	I complete the whole question before hypothesisng the answer	Do you complete the whole question before hypothesising the answer	Deductive reasoning
7	How I feel in the exam affects the way I think about the question	Does the way that you feel in the exam affect the the way that you think about the question	Student centred context
8	I come up with hypotheses immediately then change them as the question develops	Do you come up with hypotheses initially then change them as the question develops	Early hypothesis generation
9	I find that making decisions in exams is the same as making decisions in real clinical situations	Do you think making decisions in exams is the same as making decisions in real clinical situations	Student centred context
10	I find it easier to come to a decison if I recognise the case as similar to one I have seen in practice	Do you think it is easier to come to a decision if you recognise the case as similar to one you have seen before	Pattern recognition
11	I find the lack of real patients means I look for patterns as I cannot explore information	Do you find the lack of real patients means that you look for patterns as you cannot explore information	Pattern recognition
12	I find it easier to make decisions in exams as there is no emotion involved	Do you find it eaiser to make decisions in exams because there is no emotion involved	Student centred context
13	I make decisions on diagnosis or management only after completing the entire question	Do you make decisions on diagnosis or management only after completing the entire question	Deductive reasoning
14	I mostly trust my instincts when answering questions	Do you mostly trust your instincts when answering questions	Pattern recognition
15	When answering exam questions I am always looking for more unusual answers than I would in normal clinical practice	When answering exam questions do you look for more unusual answers than you would in normal clinical practice	Student centred context
16	I try to think about whether I am making inappropriate assumptions when answering questions (Biases)	Do you try to think whether you are making inappropriate assumptions when answering questions (biases)	Biases
17	If I end up with multiple hypotheses for the question answer I scrap them all and start again	If you end up with multiple hypotheses for the question answer do you scrap them all and start again	Metacognition
18	I commonly use learned mental shortcuts to help me answer questions in exams	Do you commonly use learned mental shortcuts to help me answer questions in exams	Heuristics
19	Mostly my first thought on the answer is the one that I use	Do you find that mostly your first thought on the answer is the ones that you use	Early hypothesis generation
20	I think differently in assessments compared to real clinical situations as I know only relevant information is presented	Do you think that you think differently in assessments compared to real clinical situations as you know only relevant information is presented	Student centred context
21	If I am in a rush I trust my instincts more than if I have more time	If in a rush do you trust your instincts more than if you have more time	Student centred context
22			
23	I take a patient-centred approach	Do you take a patient-centred approach?	patient centred context
24	I look for recognisable patterns within the presentation / scenario	Do you look for recognisable patterns in the scenario?	pattern recognition

[redacted]

Master spreadsheet from initial question generation showing the questions pulled in from the individual participant tabs (see along the bottom)



Copy 1 of NGT Initial Question Spreadsheet - Define questions to establish how students undertake the process of reasoning during assessments

File Edit View Insert Format Data Tools Extensions Help Last edit was on 1 February

Share

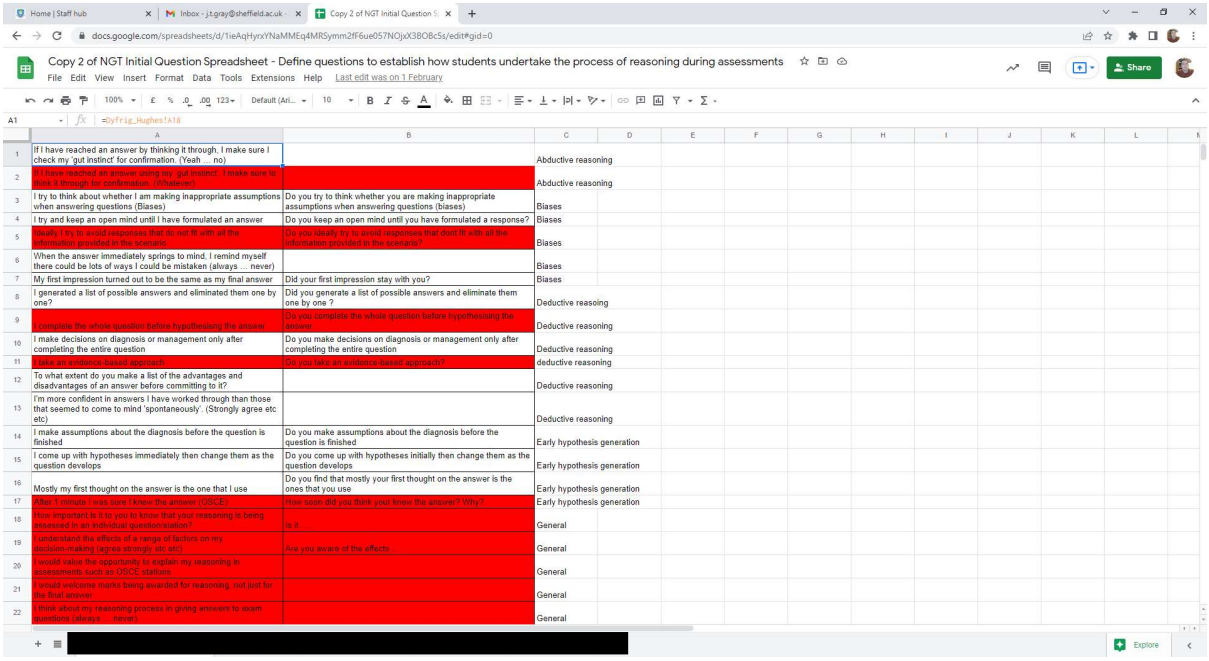
docs.google.com/spreadsheets/d/146o4Ce7qjo9mFgzNUDrfXakbfLutkHCtVki_V2qaN-Q/edit#gid=403627447

A1 Quantitative Question Form

	A	B
1	Quantitative Question Form	Qualitative Question Form
2	[redacted]	
3	I generated a list of possible answers and eliminated them one by one?	Did you generate a list of possible answers and eliminate them one by one ?
4	I was sure I knew the answer after reading the question/scenario? (SBA)	How soon did you think your knew the answer? Why?
5	After 1 minute I was sure I knew the answer (OSCE)	How soon did you think your knew the answer? Why?
6	Once I had a good idea of the answer I checked if it could be wrong	Once you had a good idea of the answer did you check if it could be wrong?
7	My first impression turned out to be the same as my final answer	Did your first impression stay with you?
8	Not all the information fitted together but I'm happy with my answer	How well did all the information fit together?
9		How did you deal with information that didn't fit your answer/hypothesis
10	The patient demographics affected my answer	How did patient demographics affect your answer
11	I had seen/observed a patient like the one in the question	Have you seen or observed a patient like the one in the question
12	At the end of the question I did check that I was happy with my answer	At the end of the question did you check your answer with yourself
13	My nervousness about the exam affected my answer	Did your nervousness about the exam affect your answer?
14	I used a systematic list or mnemonic to help me	Did you use any systematic list or mnemonic to help you?
15	I was keen to rule in or out the most serious possibility	How keen were you to rule in or out the most serious condition
16	I find descion making in exams is the same as "real life"	To what extent do you find decsion making in exams is like "real life" clinical decsion making
17	I find it hard to "suspend disbelief" in exams	To what extent do you find it hard to "suspend disbelief" in exams
18	I expect there to be more rare diseases in exams	Do you expect there to be more rare diseases in exams? If so how does this affect your decision making
19	I was concerned my answer was too obvious	Were you concerned your answer was too obvious? What did you do to check this out?
20	I was concerned my answer was too obscure	Were you concerned your answer was too obscure? What did you do to check this out?
21	The patient/simulators responses affected my hypotheses	in what way did the patient/simulators responses affect your hypotheses
22	The patient/simulators responses affected my final answer	in what way did the patient/simulators responses affect your final answer ?
23		
24		
25		
26		
27		

[redacted]

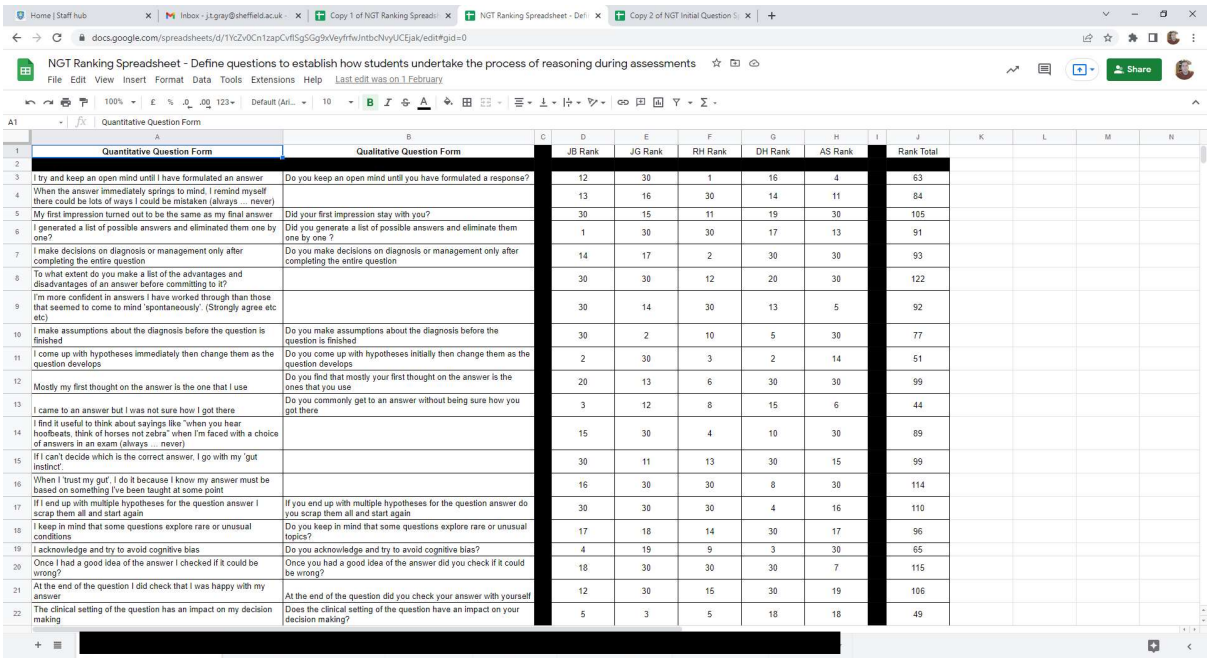
Example of an individual's question generation tab which feeds into the above master spreadsheet.



Copy 2 of NGT Initial Question Spreadsheet - Define questions to establish how students undertake the process of reasoning during assessments

1	If I have reached an answer by thinking it through, I make sure I check my 'gut instinct' for confirmation. (Yeah … no)		Abductive reasoning
2	If I have reached an answer using my 'gut instinct', I make sure to think it through for confirmation. (Whatever)		Abductive reasoning
3	I try to think about whether I am making inappropriate assumptions when answering questions (Biases)	Do you try to think whether you are making inappropriate assumptions when answering questions (biases)	Biases
4	I try and keep an open mind until I have formulated an answer	Do you keep an open mind until you have formulated a response?	Biases
5	Ideally I try to avoid responses that do not fit with all the information provided in the scenario	Do you ideally try to avoid responses that dont fit with all the information provided in the scenario?	Biases
6	When the answer immediately springs to mind, I remind myself there could be lots of ways I could be mistaken (always … never)		Biases
7	My first impression turned out to be the same as my final answer	Did your first impression stay with you?	Biases
8	I generated a list of possible answers and eliminated them one by one?	Did you generate a list of possible answers and eliminate them one by one ?	Deductive reasoing
9	I complete the whole question before hypothesising the answer	Do you complete the whole question before hypothesising the answer	Deductive reasoning
10	I make decisions on diagnosis or management only after completing the entire question	Do you make decisions on diagnosis or management only after completing the entire question	Deductive reasoning
11	I take an evidence-based approach	Do you take an evidence-based approach?	deductive reasoning
12	To what extent do you make a list of the advantages and disadvantages of an answer before committing to it?		Deductive reasoning
13	I'm more confident in answers I have worked through than those that seemed to come to mind 'spontaneously'. (Strongly agree etc etc)		Deductive reasoning
14	I make assumptions about the diagnosis before the question is finished	Do you make assumptions about the diagnosis before the question is finished	Early hypothesis generation
15	I come up with hypotheses immediately then change them as the question develops	Do you come up with hypotheses initially then change them as the question develops	Early hypothesis generation
16	Mostly my first thought on the answer is the one that I use	Do you find that mostly your first thought on the answer is the ones that you use	Early hypothesis generation
17	After 1 minute I was sure I knew the answer (OSCE)	How soon did you think your knew the answer? Why?	Early hypothesis generation
18	How important is it to you to know that your reasoning is being assessed in an individual question/station?	is it	General
19	I understand the effects of a range of factors on my decision-making (agree strongly etc etc)	Are you aware of the effects	General
20	I would value the opportunity to explain my reasoning in assessments such as OSCE stations		General
21	I would welcome marks being awarded for reasoning, not just for the final answer		General
22	I think about my reasoning process in giving answers to exam questions (always … never)		General

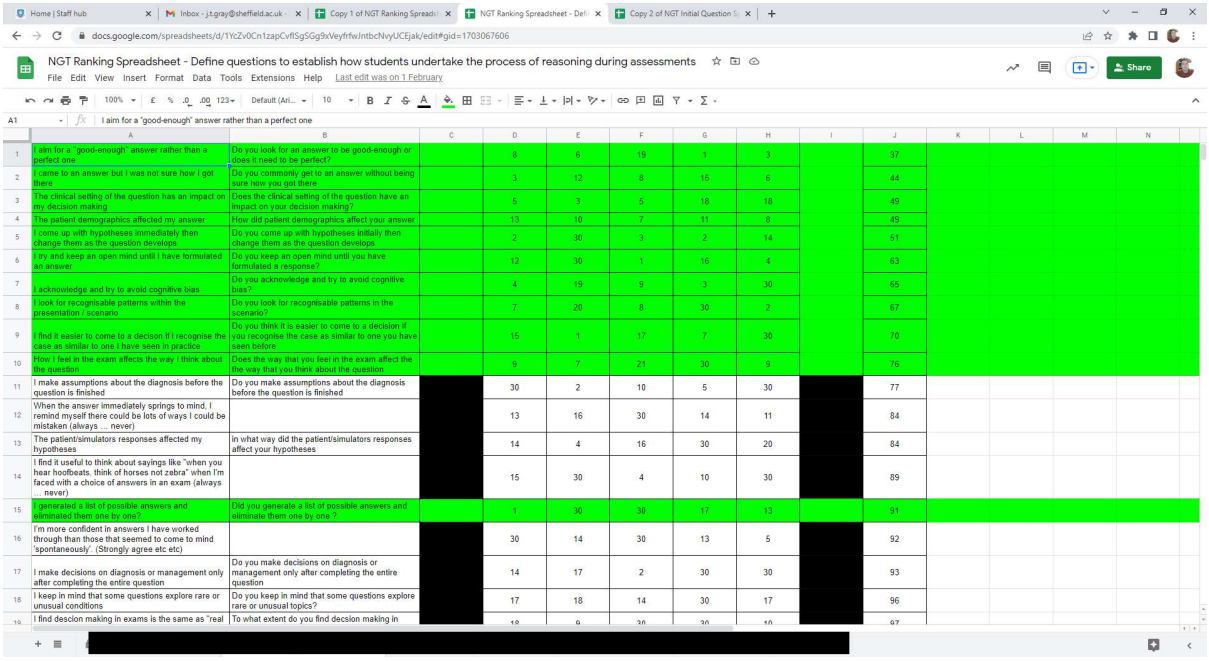
Further development of the initial question spreadsheet showing its development during the discussion phase. Those in red were considered unhelpful or duplicate and so removed at this stage.



NGT Ranking Spreadsheet - Define questions to establish how students undertake the process of reasoning during assessments

	Quantitative Question Form	Qualitative Question Form	JB Rank	JG Rank	RH Rank	DH Rank	AS Rank	Rank Total
3	I try and keep an open mind until I have formulated an answer	Do you keep an open mind until you have formulated a response?	12	30	1	16	4	63
4	When the answer immediately springs to mind, I remind myself there could be lots of ways I could be mistaken (always … never)		13	16	30	14	11	84
5	My first impression turned out to be the same as my final answer	Did your first impression stay with you?	30	15	11	19	30	105
6	I generated a list of possible answers and eliminated them one by one?	Did you generate a list of possible answers and eliminate them one by one ?	1	30	30	17	13	91
7	I make decisions on diagnosis or management only after completing the entire question	Do you make decisions on diagnosis or management only after completing the entire question	14	17	2	30	30	93
8	To what extent do you make a list of the advantages and disadvantages of an answer before committing to it?		30	30	12	20	30	122
9	I'm more confident in answers I have worked through than those that seemed to come to mind 'spontaneously'. (Strongly agree etc etc)		30	14	30	13	5	92
10	I make assumptions about the diagnosis before the question is finished	Do you make assumptions about the diagnosis before the question is finished	30	2	10	5	30	77
11	I come up with hypotheses immediately then change them as the question develops	Do you come up with hypotheses initially then change them as the question develops	2	30	3	2	14	51
12	Mostly my first thought on the answer is the one that I use	Do you find that mostly your first thought on the answer is the ones that you use	20	13	6	30	30	99
13	I came to an answer but I was not sure how I got there	Do you commonly get to an answer without being sure how you got there	3	12	8	15	6	44
14	I find it useful to think about sayings like "when you hear hoofbeats, think of horses not zebra" when I'm faced with a choice of answers in an exam (always … never)		15	30	4	10	30	89
15	If I can't decide which is the correct answer, I go with my 'gut instinct'.		30	11	13	30	15	99
16	When I 'trust my gut', I do it because I know my answer must be based on something I've been taught at some point		16	30	30	8	30	114
17	If I end up with multiple hypotheses for the question answer I scrap them all and start again	If you end up with multiple hypotheses for the question answer do you scrap them all and start again	30	30	30	4	16	110
18	I keep in mind that some questions explore rare or unusual conditions	Do you keep in mind that some questions explore rare or unusual topics?	17	18	14	30	17	96
19	I acknowledge and try to avoid cognitive bias	Do you acknowledge and try to avoid cognitive bias?	4	19	9	3	30	65
20	Once I had a good idea of the answer I checked if it could be wrong?	Once you had a good idea of the answer did you check if it could be wrong?	18	30	30	30	7	115
21	At the end of the question I did check that I was happy with my answer	At the end of the question did you check your answer with yourself	12	30	15	30	19	106
22	The clinical setting of the question has an impact on my decision making	Does the clinical setting of the question have an impact on your decision making?	5	3	5	18	18	49

Master spreadsheet for the question ranking process. Questions were pulled through from the original spreadsheet and duplicated in each individual tab to allow each member to rank them individually without influence. These then populated and scored in the master sheet. Individual tabs can be seen along the bottom.



NGT Ranking Spreadsheet - Define questions to establish how students undertake the process of reasoning during assessments

1	I aim for a "good-enough" answer rather than a perfect one	Do you look for an answer to be good-enough or does it need to be perfect?		8	6	19	1	3		37					
2	I came to an answer but I was not sure how I got there	Do you commonly get to an answer without being sure how you got there		3	12	8	15	6		44					
3	The clinical setting of the question has an impact on my decision making	Does the clinical setting of the question have an impact on your decision making?		5	3	5	18	18		49					
4	The patient demographics affected my answer	How did patient demographics affect your answer		13	10	7	11	8		49					
5	I come up with hypotheses immediately then change them as the question develops	Do you come up with hypotheses initially then change them as the question develops		2	30	3	2	14		51					
6	I try and keep an open mind until I have formulated an answer	Do you keep an open mind until you have formulated a response?		12	30	1	16	4		63					
7	I acknowledge and try to avoid cognitive bias	Do you acknowledge and try to avoid cognitive bias?		4	19	9	3	30		65					
8	I look for recognisable patterns within the presentation / scenario	Do you look for recognisable patterns in the scenario?		7	20	8	30	2		67					
9	I find it easier to come to a decison if I recognise the case as similar to one I have seen in practice	Do you think it is easier to come to a decision if you recognise the case as similar to one you have seen before		15	1	17	7	30		70					
10	How I feel in the exam affects the way I think about the question	Does the way that you feel in the exam affect the the way that you think about the question		9	7	21	30	9		76					
11	I make assumptions about the diagnosis before the question is finished	Do you make assumptions about the diagnosis before the question is finished		30	2	10	5	30		77					
12	When the answer immediately springs to mind, I remind myself there could be lots of ways I could be mistaken (always ... never)			13	16	30	14	11		84					
13	The patient/simulators responses affected my hypotheses	in what way did the patient/simulators responses affect your hypotheses		14	4	16	30	20		84					
14	I find it useful to think about sayings like "when you hear hoofbeats, think of horses not zebra" when I'm faced with a choice of answers in an exam (always ... never)			15	30	4	10	30		89					
15	I generated a list of possible answers and eliminated them one by one?	Did you generate a list of possible answers and eliminate them one by one ?		1	30	30	17	13		91					
16	I'm more confident in answers I have worked through than those that seemed to come to mind 'spontaneously' (Strongly agree etc etc)			30	14	30	13	5		92					
17	I make decisions on diagnosis or management only after completing the entire question	Do you make decisions on diagnosis or management only after completing the entire question		14	17	2	30	30		93					
18	I keep in mind that some questions explore rare or unusual conditions	Do you keep in mind that some questions explore rare or unusual topics?		17	18	14	30	17		96					
19	I find desclon making in exams is the same as "real	To what extent do you find decsion making in		18	9	30	30	10		97					

This shows the final post ranking list re-ordered to put into order. The use of the question lower down was it was felt that 10 questions was ideal, and this was the next highest to provide additional information rather than simply revisiting existing topics in the top 9 questions.

Appendix 7 – Example Single Best Answer Questions used in the study

A 52 year old Afro-Caribbean man attends General Practice following a routine health screening at work. He was found to be hypertensive. She has done some home blood pressure readings which average 160/92 mmHg. He is asthmatic and has occasional issues of urinary urgency.

Examination is normal. Urine dip testing is negative.

Which is the most appropriate anti-hypertensive to initiate treatment with?

A. ACE inhibitor
B. Alpha blocker
C. Beta-blocker
D. Calcium channel blocker
E. Thiazide-like diuretic

Correct answer D – Calcium channel blocker

NICE guidance is all those of Afro-Caribbean origin should be offered a CCB as a first line treatment for hypertension. The others are all valid BP treatments however B-blockers are contraindicated due to the asthma. E may lead to more urinary issues and, on the other side B may be a second line option if prostate issues but not first line. A is a second line option in this patient.

A six year old child is brought to the Emergency Department. Mum has noticed a limp over the past few months, but it has got worse since they were playing with friends in the park. The child has complained of stiffness in the morning for the past two months in the hip but also sometimes the elbows and knees.

On examination, temperature 36.8 °C, BMI 26.4 kg/m^2. The hip is uncomfortable to move but without restriction.

Which is the most likely diagnosis?

A. Irritable hip
B. Juvenile arthritis
C. Perthes disease
D. Slipped femoral epiphysis
E. Trauma

Correct answer B – Juvenile arthritis

The chronicity of the pain and age makes this most likely. SUFE and trauma are unlikely to have the chronicity. Perthes could but the multiple joint involvement makes this less likely.

A 48 year old woman attends General Practice with a one week history of facial pain. She describes sudden attacks of sharp pain on the left side of the scalp and forehead running to the front. She describes it as 9/10 in severity and it sometimes causes her eye to water and associated nasal congestion. Each attack lasts a few minutes and can be provoked by eating. She has not had this before.

On examination, vital signs are normal. There is scalp tenderness all across the upper left scalp and face. There is no rash evident.

Which is the most appropriate first line treatment for the likely cause?

A. Aciclovir
B. Carbamazepine
C. Intranasal steroids
D. Intranasal sumatriptan
E. Oral steroids

Correct answer B – Carbamazepine

The description is of trigeminal neuralgia with typical attack pattern. Lacrimation and nasal congestion can be associated as can the scalp tenderness. The eating provocation makes GCA less likely and also sinusitis. With no rash shingles is also unlikely. Carbamazepine is first line for TN.

A 22 year old woman presents to the Emergency Department with pelvic pain. The pain started eight hours ago and has increased since felt most in the right iliac fossa. She has noticed some vaginal blood spotting but no discharge. Her last menstrual period was 7 weeks ago but she reports these as usually erratic and widely spaced. She has a new partner, and they are not using contraception.

On examination, temperature 36.8 °C, pulse 104 bpm, blood pressure 110/64 mmHg. There is right iliac fossa tenderness on palpation. Vaginal examination shows no cervical excitation.

Which is the most likely diagnosis?

A. Appendicitis
B. Ectopic pregnancy
C. Miscarriage
D. Pelvic inflammatory disease
E. Ruptured ovarian cyst

Correct answer B – Ectopic pregnancy

There is a high risk of pregnancy in this case. Miscarriage is less likely with the location of the pain and spotting rather than bleeding. PID is a potential, especially given age and sexual history, but there is no cervical excitation or pyrexia. Ovarian cyst rupture will not give PV bleeding. Appendicitis is possible but with the PV blood a gynae cause is more likely.

A 23 year old man presents to General Practice with a two year history of recurrent abdominal pain The pain occurs once or twice a week and is associated with very loose, and more frequent, bowel motions. He has lost weight over the last couple of months and feels the symptoms are worsening. He has not noticed any blood in the motions and is unsure of any link to specific foods although bread tends to make him feel bloated.

Examination is normal.

Which is the most likely diagnosis?

A. Bowel cancer
B. Coeliac disease
C. Crohn's disease
D. Diverticulitis
E. Irritable bowel syndrome

Correct answer E - Irritable bowel syndrome.

Pt is young so cancer and diverticulitis are unlikely. The symptoms fit the ROME IV criteria for diagnosis. Crohn's is possible but the history is not as typical. Bloating with bread is insufficient to suggest coeliac as the cause.

A 46 year old woman is brought to the Emergency Department following an industrial accident. During the accident her left thigh was crushed by falling concrete and only released after around 45 minutes. She is complaining of severe pain in the left thigh.

On examination, pulse 54 bpm regular and there are reduced tendon reflexes on neurological examination.

ECG confirms the bradycardia and shows indistinct p-waves, prolonged P-R interval and widening of the QRS interval.

Which is the most appropriate initial treatment?

A. Bisphosphonates
B. Calcium gluconate
C. Haemodialysis
D. Insulin
E. Potassium replacement

Answer B - Calcium gluconate

This is typical of hyperkalaemia secondary to a crush injury. ECG changes indicate potentially life threatening so IV calcium gluconate is the treatment of choice. Insulin and haemodialysis are later options. Bisphosphonates are for hypercalcaemia and potassium replacement for hypokalaemia

A 50 year old patient attends General Practice with a three year history of tinnitus. They describe a constant buzzing sound in the left ear that has been getting progressively worse. They have also noticed progressive hearing loss in the left ear. More recently they have been getting episodes of dizziness which they describe as “like the room spinning”. These occur randomly and not specifically related to certain head movements. They have also had some recent headaches.

Examination of the ear canals is unremarkable.

Which is the most likely diagnosis?

A. Acoustic neuroma
B. Benign paroxysmal vertigo
C. Cholesteatoma
D. Labyrinthitis
E. Meniere’s disease

Answer A – Acoustic neuroma

The combination of tinnitus, hearing loss and vertigo puts AN and Meniere’s front and centre for the differential however Meniere’s tend to be episodic whilst this is persistent and progressive making An more likely. BPV and labyrinthitis are unlikely to have the triumvirate of symptoms and Cholesteatoma is likely to be visible in the ear canal.

A 64 year old man presents to the Emergency Department having noticed fresh red blood in his urine over the past 24 hours. It is accompanied by pain and irritation on passing urine. 3 days ago he underwent cystoscopy due to a raised PSA of 4.5 µg/l and is awaiting the results but was told that it looked ok. He reports reasonable urinary stream and no urethral discharge.

On examination he is apyrexial, pulse 76 bpm, blood pressure 132/82 mmHg.

Urine dip shows Blood +++, protein + and white cells ++. Nitrates are negative.

Which is the most likely diagnosis?

A. Bladder cancer
B. Non-specific urethritis
C. Prostate cancer
D. Ureteric stone
E. Urinary tract infection

Answer E – Urinary tract infection

The combination of blood with dysuria and the recent cystoscopy makes a UTI most likely. The lack of urethral discharge means that NSU is unlikely. Although the PSA is slightly raised the lack of anything visible makes prostate cancer less likely as the bleed source. There is insufficient information to suggest a stone or bladder cancer.

A 50 year old woman presents to General Practice with numbness in the feet and lower legs. She has noticed this coming on slowly over the past few years, but it is now affecting her walking. She is also feeling tired all the time. She says that she had a bad chest infection a few years ago and wonders if this has come on since then. She has no other past medical history. Her twin sister is diabetic. She is a smoker of 35 pack years and has a vegan diet. She has no past medical history, but her twin sister is diabetic.

On examination, vital signs are normal. She looks a little pale. There is a stocking distribution sensory deficit.

Which test is most likely to confirm the underlying cause of the symptoms?

A. Ankle-brachial pressure index
B. Haemoglobin A1c
C. Nerve conduction studies
D. Thyroid function test
E. Vitamin B12 levels

Answer E – Vitamin B12 levels

This woman is vegan which immediately makes B12 deficiency higher risk. The slow onset fits with this as does the fatigue – likely secondary to anaemia (with the pallor). Diabetes is a possibility esp. with the family history but no symptoms are described, likewise PVD is possible but there is no leg pain described. Nerve conduction studies will not show an underlying cause

A 63 year old patient presents to the eye casualty with sudden loss of vision in the right eye. They have noticed sensitivity to light, some visual disturbance and eye discomfort over the past few days with an associated mild headache. Today they had sudden onset of eye pain with visual loss on the right side.

On examination the right eye is red, the pupil is fixed and dilated, and the red reflex is present.

Which is the most likely diagnosis?

A. Acute glaucoma
B. Migraine
C. Retinal artery occlusion
D. Stroke
E. Vitreous haemorrhage

Answer A – Acute glaucoma

This describes an episode of acute closed angle glaucoma; Migraine is unlikely to cause sudden visual loss and will not cause the pupil changes. Vitreous haemorrhage would have no red reflex. Retinal artery occlusion would have the fixed pupil but not a painful eye and nor would a CVA.

Appendix 8 – Example OSCE Station used in the study

Research OSCE Station 1

Diagnosis: Abdominal Pain

Simulated male/female (White) patient 20-30 years old.

EQUIPMENT LIST

Hand sanitiser

Chairs for examiner, patient, and student

INSTRUCTIONS FOR THE STUDENT

On this station you will be examined on communication skills as well as your ability to formulate a diagnosis and management of this.

You have XXXXX minutes to complete the task above. The examiner will then ask you two questions relating to the case.

Diagnosis: Abdominal Pain

INSTRUCTIONS FOR THE PATIENT / ACTOR

Please answer any questions based on the following scenario.

The student will have XXXXX minutes for this discussion.

You are Paul/Paula Smith 25 years old (d.o.b. 13th January 1998) and have attended the practice due to a three-month history of abdominal pain and change in bowel habit.

Over the past three months you have noticed that you have been getting a change to your bowel habit. This is usually episodes of diarrhoea (looser motions and going up to six or seven times daily) although you have had occasional constipation. You have been getting some cramping abdominal pain in the lower stomach which seems to be relieved with defaecation.

WHEN ASKED:

You have not noticed any blood in the motions but do occasionally get blood on wiping with some anal itching. This has been an issue for many years and worse after occasional constipation.

You have lost a little weight (around 4kg) over the past 3 months but have been actively seeking to do so as feel it would be good for your health. You have not noticed any other general physical symptoms.

There is a family history of bowel cancer – your grandfather died two weeks ago of this. You are worried that you might have the same thing. There is no other history of bowel issues.

You have been wondering if you might be a little depressed as you have felt tearful at work. This has been for around the same time period as the change in the bowels.

You take no medication routinely but when the pain has been bad you have tried some paracetamol which has not really helped.

You have no allergies.

You smoke.10 cigarettes a day though keep trying to stop. When you stop smoking your bowels seem a little worse with more diarrhoea. You drink around 8 pints of beer a week. You work in a GP surgery as a receptionist, and it has been very stressful as you are understaffed and are having to work more than you had hoped to cover. The symptoms do seem worse when you are feeling stressed.

You have noticed that sometimes spicy or rich food seems to make the symptoms worse. You have not noticed any link between the discomfort and eating bread or pasta.

You want a scan to try to work out what's wrong with your stomach.

Following the completion of the Station You will be asked to grade the student's consultation skills as follows:

X marks = Excellent listening and use of language, fluent, clear explanations, empathic and highly supportive

X marks = Good listening skills, mostly clear language, fluent, some gaps in explanation, empathetic and quite supportive

X mark = Mostly clear language but poor listening skills, fragmented explanation, poor empathy and minimal support

(Note: students in this category may be asking pre-rehearsed questions by rote rather than exploring the symptoms that you have told them about in more depth).

X marks = Poor listening skills and unclear language, incoherent explanation, no empathy

Listening skills: Did the student make eye contact, encourage you to talk by using open questions and gestures, summarise what you have told them and respond appropriately to what you have told them?

Language and explanation: Did they ask questions and explain things in a way that you found easy to understand, without using medical jargon?

Support and empathy: Did you feel supported, and that the student understood your feelings (empathy)?

INSTRUCTIONS FOR THE EXAMINER

The student has been given the following instructions:

Please consult with this patient who attends General Practice to discuss abdominal pain.

On this station you will be examined on communication skills as well as your ability to formulate a diagnosis and management of this.

You have XXXXX minutes to complete the task above. The examiner will then ask you two questions relating to the case.

Please observe the student closely and mark them according to the marking grid provided.

You must also provide a global rating.

Usually, there will be a close correlation between the marks and the global rating. Sometimes a student will achieve a good score on their marks, but you gained a clear impression that they did not show deep understanding. You might then give them a lower global rating. There should not be a very wide difference between the two. Above all, please be consistent between students.

At XXXX minutes, but not before, please ask the following questions:

Do not deviate from the script by asking different or additional questions.

Question 1: What do you think is the most likely diagnosis?

Answer: Irritable bowel syndrome (1 mark)

Question 2: Give two appropriate ways to help manage this patient's symptoms?

Answer (Any 2 of): Dietary Advice, Anti-spasmodic medication, Stress management, Good fluid intake, regular physical activity.

Once the student has left the station, ask the patient to award their marks out of 3 (see instructions to simulated patients) and include these on your mark sheet. These 3 marks are entirely at the discretion of the simulated patient.

EXAMINER'S MARK SHEET

Student's OSCE No .....

Please write the student's OSCE number on the top of this mark sheet.

Details of Task	Performs skills competently and to a high standard	Performs almost completely and to a good standard	Performs most of skill to reasonable standard	Performs part of skill or to a poor standard	Performs very poorly
Identifies self and patient. Establishes co-operation and rapport.	1	1	1	0	0
Takes a clear history, asking open questions and gathering information around the change in bowel habit	4	3	2	1	0
Confirms the presence or absence of red flag symptoms - FH, weight loss/ pain/ progression/ systemic symptoms	5	4	2	1	0
Takes social history, including lifestyle factors, work, exercise, diet etc.	4	3	2	0	0
Reviews past medical history, medications, allergies	1	1	1	0	0
Establishes patients concerns and expectations	2	1	1	0	0
Good communication skills with logical question approach and good rapport and engagement	2	1	1	0	0

Answers question 1 correctly	1	1	1	0	0
Answers question 2 correctly	2	1	1	0	0
Marks awarded by patient (see patient's instructions for details)	3	2	2	1	0
Total marks	**25 marks**				
Before submitting this form, please check that you have entered a mark on every line of this sheet					

GLOBAL RATING (Please circle)	EXCELLENT PASS	GOOD PASS	PASS	BORDERLINE	CLEAR FAIL

Examiner's name: (block capitals) .....

Examiner's signature: .....

Students get a great deal of benefit from good feedback.

Please use the time between students to provide feedback. Please make sure it is personal, constructive and useful. Feedback is MANDATORY if you give a borderline or fail rating, and desirable for all students.

The feedback you give is provided to the students after they get their results.

Appendix 9 – Ethics approval for the primary study

Downloaded: 14/05/2022
Approved: 10/05/2022

James Gray
Registration number: 200278917
Medical School
Programme: FCM MDH Postgraduate Research Programme

Dear James

PROJECT TITLE: Understanding how healthcare students undertake clinical reasoning in examinations - Part 2, Mixed Methods Study to determine student reasoning utilising questions from part 1
APPLICATION: Reference Number 044729

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 10/05/2022 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 044729 (form submission date: 13/04/2022); (expected project end date: 31/08/2025).
- Participant information sheet 1100639 version 2 (13/04/2022).
- Participant information sheet 1100640 version 2 (13/04/2022).
- Participant consent form 1100643 version 2 (13/04/2022).
- Participant consent form 1100642 version 2 (13/04/2022).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Charlotte Cole
Ethics Administrator
Medical School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Appendix 10 – Email correspondence regarding use of pixel record

University of Sheffield

James T Gray <j.t.gray@sheffield.ac.uk>

Pixel Recorder **URGENT**

[redacted] 26 February 2023 at 18:56
To: Michelle Marshall <m.marshall@sheffield.ac.uk>
Cc: James T Gray <j.t.gray@sheffield.ac.uk>

Hi James,

I own a Pixel 7 Pro and the 'recorder' app function is fantastic. It relies on AI onboard the device, so no information is exchanged or processed in the cloud - where we would have concerns if this was the case. As long as your device is secured by password/PIN and only used for work purposes - i.e. only your sheffield.ac.uk google account is in use then this workflow for interviews is actually more secure than an encrypted digital recorder used in other areas of the Faculty e.g. ScHARR. I have absolutely no problem with you using the Pixel for interviews - let me know if anyone has suggested different. My only comment re good working practice is that you may want to transfer interviews, transcripts and any other other data from your google drive to a filestore folder (X:Drive) dedicated to your project. This lessens the risk of sharing by mistake but itself is a small risk if using your own private UoS Google Drive areas.

Of course, usual ethics, governance and GDPR rules apply in terms of consent and making sure interviewees are aware of recording, transcribing et. You have probably undertaken a Data Protection Impact Assessment already as part of your project's ethics etc. File names and transcription texts should be suitably semi-anonymised according to the project's research data management plan. No need to tell you all this of course.

Let me know if I can help further.

Best Wishes,

[redacted]

[redacted]

IT Services, The University of Sheffield

10-12 Brunswick Street, Sheffield S10 2FN

T: 0114 222 8766

M: [redacted]
E: [redacted]

Google Chat – Preferred for direct messaging
https://www.sheffield.ac.uk/it-services

Appendix 11 – Code map

Primary Themes	Subcode	Subcode (Level 2)
Abductive Reasoning	Flexibility in thinking	
	Variability based on available data	
Assessment Centred Context	Assessment v Real World	
	OSCE Specific	Effect of stress
		Examiner presence
		Mark Scheme Approach
		Patient script related thinking
		OSCE Setting
		Station order
		Task Specific Thinking
		Thinking development through the exam
		Time driven behaviours
	SBA Specific	Effect o Stress
		Environment for the Examination
		Question order
		Question writer link
		Task specific thinking
		Thinking Development through the exam
		Time Pressure
	SPLD Implications	
	Summative v Formative	
Biases	Anchoring Bias	
	Availability bias	
	Confirmation Bias	
	Gamblers fallacy	
	Premature Closure	
	Representativeness Bias	
Deductive reasoning	Use of rule out in SBA	
	Advantage in OSCE	
Early Hypothesis Generation	Based on past experience	
	Difference in exam types	
	Difference in decision type	
	Issues when unable to hypothesise	
	Link to early closure	
	Use of DD as early hypo. In OSCE	
	Use of SBA answers as a proxy	
Metacognition	Control of biases	
	Development with experience	
	Effects of Stress	
	Exam process management	
Patient Centred Context	Clinical Setting	Assumption of Severity
		Change in expectation
		Difference in exam types
		Limited Impact in examinations
		Relevance in OSCE
		Relevance to part of decision
	Demographics	Difficulty if mismatches with hypothesis
		Effect on management decisions
		Dominance in thinking
		Failure to Acknowledge
		Impact of ethnicity
		Link to differential diagnosis
		Links to real patient in OSCE
		SBA v OSCE Difference
	Lack of Correlation in OSCE	
Pattern Recognition	Assessment type driving it	
	Challenge of contradictory information	
	Fundamental reality of job or life	
	OSCE related patterns	
	Recognition but not clear why	
	Regional Variation	
	Related to learning approach	
	Requirement for question writing	
	Risks of	
	Trigger words	
Satisficing	Answers feeling right	
	Exam Differences	
	Linked to question difficulty	
	Pressure Driven	
	Red Flags	
	Student perceptions of	
Student Centred Context	Clinical role	
	Different Student Reasoning Styles	
	Personal affect in the assessment	
	Topic Confidence	
	Values and Beliefs	

	Pilot (Top Level Codes)	15
	Group 1	34
	Group 2 (	16
	Group 3	7
	Group 4	4
	Group 5	4
	Group 6	4
	Group 7	4
	Top level Code review	1
		89

Appendix 12 – COREQ Checklist

Topic	Item No.	Guide Questions/Description	Reported in Section
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	5.5.2/8.1
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	1.2
Occupation	3	What was their occupation at the time of the study?	1.2
Gender	4	Was the researcher male or female?	N/R
Experience and training	5	What experience or training did the researcher have?	1.2
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	1.2/11.7
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	6.3.1
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons, and interests in the research topic	6.3.1
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	3.4
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	5.6.1 (Pilot), 6.2
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	6.2
Sample size	12	How many participants were in the study?	6.2.3
Non-participation	13	How many people refused to participate or dropped out? Reasons?	N/A
Setting			
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	6.4/6.5
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	6.5
Description of sample	16	What are the important characteristics of the sample? e.g. demographic, data, date	7.2/8.1
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	4.2/5.2
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	N/A
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	5.5.1/5.5.2
Field notes	20	Were field notes made during and/or after the interview or focus group?	6.5

Duration	21	What was the duration of the inter views or focus group?	8.1/8.2
Data saturation	22	Was data saturation discussed?	3.4.3
Transcripts returned	23	Were transcripts returned to participants for comment and/or	N/A
Domain 3: analysis and findings			
Data analysis			
Number of data coders	24	How many data coders coded the data?	5.6.3/6.7
Description of the coding tree	25	Did authors provide a description of the coding tree?	App 9
Derivation of themes	26	Were themes identified in advance or derived from the data?	4.2.7
Software	27	What software, if applicable, was used to manage the data?	6.7
Participant checking	28	Did participants provide feedback on the findings?	N/A
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Ch 8 & 9
Data and findings consistent	30	Was there consistency between the data presented and the findings?	Ch 11
Clarity of major themes	31	Were major themes clearly presented in the findings?	Ch 10
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	Ch 10

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Appendix 13 - Glossary

Abduction: a form of logical inference formulated and advanced by Charles Sanders Peirce seeking to find the most logical conclusion from a set of data. It leads to a plausible solution but does not positively verify it.

Analytical: Decision-making which involves reviewing data and ascertaining the patterns within it.

Bayesian logic: Also known as probabilistic decision-making. This applies probability theory to inductive and abductive processes.

Bounded rationality: A concept that rationality is limited when we make decisions, due to problem difficulty, cognitive capacity and time available.

Categorical: A set of appropriate rules and processes that apply to specific clinical situations.

Cognitive biases: A predictable pattern of deviation from normal rational judgement

Cognitive continuum theory: This suggests that rather than a very binary dual-process approach to problem solving we instead adopt an approach on a continuum which has analytical at one end and non-analytical at the other.

Deduction: The process of taking data and reasoning to a logical conclusion which can be verified.

Dual-process theory: A concept of cognition which suggests that when faced with a problem we either utilise a rapid non-analytical or slow analytical process to address it.

Evidence based medicine: The explicit use of the most up-to-date evidence applied to facilitate decision-making in individual patient care.

Heuristics: An approach to problem solving that is not optimal but sufficient to reach a goal. Usually based on previous experience.

Hypothesis: An assumption made on evaluation of a situation that can then be tested by data.

Hypothetico-deductive method: Taking a hypothesis and testing it through data gathering and evaluation.

Induction: The use of experience and observation to produce a reasonable explanation.

Instrumental: Decision-making that focusses on a necessary and effective approach to meet a set goal.

Medical Model: Medical model is the term for the "set of procedures in which all doctors are trained" (Laing, 1971). It includes complaint, history, physical examination, ancillary tests if needed, diagnosis, treatment, and prognosis with and without treatment. The medical model embodies basic assumptions about medicine that drive research and theorizing about physical or psychological difficulties on a basis of causation and remediation.

Non-analytical: Decision-making which is fast and requires no data analysis.

Pattern recognition: A form of inductive or non-analytical decision-making in which immediate similarity to previous experience is recognised in presentation of data. In clinical reasoning usually considered with respect to illness scripts.

Probabilistic: See Bayesian reasoning.

Satisficing: a decision-making approach whereby a decision on a problem is made at a level of acceptability to the individual decision maker.