Validity of a Virtual Reality-Based Clinical Case for Assessment of Clinical Competence

Melyssa Claire Roy

A thesis submitted for the degree of
Master of Medical Science
at the University of Otago,
Dunedin, New Zealand.

May 2013
Abstract

Reliable and valid assessment of the clinical competence of medical students and doctors is essential for the safety of patients. Current modes of assessment are limited in their ability to evaluate some key aspects of competence, such as clinical reasoning ability and timely decision-making. The aims of this study were to assess the validity of a virtual reality-based clinical case as a method of assessment of clinical competence. In addition, this study intended to specifically examine the capacity of the virtual reality case format to assess clinical reasoning ability. The Otago Virtual Hospital is a virtual reality-based computer programme in which the performance of doctors and students can be assessed while managing a simulated clinical case in real time. As a pilot study, 12 participants, (consisting of three cohorts comprised of third-year medical students, fifth-year medical students and qualified doctors) from Otago Medical School participated in a simulated clinical case. Their performance was measured by scoring their achievement of set outcomes representing optimal management; these were developed from expert opinion. Qualitative thematic analysis of case transcripts was undertaken to compare clinical reasoning skills.

Scores of performance from the virtual reality case showed qualified doctors achieved the highest scores, significantly higher than the third-year student group. Qualified doctors were also significantly better able to make correct full diagnoses and achieve safe clinical management compared with the student cohorts. These results showed some significant differences between groups at different stages of medical training, hence supporting the construct validity of the virtual-reality based clinical case. Thematic analysis to identify clinical reasoning themes indicated that compared to the student cohorts, the qualified doctor group was better able to transform information into key clinical concepts, generate more accurate diagnoses, and generate correct diagnoses more efficiently. With increasing clinical experience there was a superior ability to communicate clinical information succinctly and precisely, and to construct effective patient management plans.
The virtual reality based clinical case provided an authentic clinical task. Within the programme, overall performance and clinical reasoning abilities could be assessed by analysis of a summary admission note at the conclusion of the case. These results suggest that simulated virtual cases may provide a valid and rapid means of assessing clinical competence, and can provide more comprehensive information about clinical reasoning ability than traditional means of assessment.
Acknowledgements

With thanks to my supervisors: Professor Tim Wilkinson, Dr Phil Blyth and Professor Rob Walker, for their time, expertise and contributions.

Also, with appreciation to the University of Otago, for the Emily Johnston Scholarship 2012.

For the invaluable advice on academic writing, thank you to Dr Jon Cornwall.

Finally, with thanks to actor and colleague, Dr Chris Hopkins.
Table of Contents

Abstract ........................................................................................................................................ ii
Acknowledgements .................................................................................................................... iv
Table of Contents ...................................................................................................................... v
List of Tables ............................................................................................................................... vii
List of Figures ............................................................................................................................. ix
Definitions ..................................................................................................................................... x

Chapter One: Introduction ......................................................................................................... 1

1.1 Overview ............................................................................................................................. 1
1.2 Clinical competence ............................................................................................................ 2
  1.2.1 Clinical reasoning ........................................................................................................ 3
1.3 Assessment of clinical competence .................................................................................... 8
  1.3.1 Specific assessment of clinical reasoning ability ....................................................... 12
1.4 Simulation in medical education ....................................................................................... 16
  1.4.1 Historical perspective ............................................................................................... 16
  1.4.2 Factors driving development of simulation .............................................................. 17
  1.4.3 The local perspective: Simulation in Australasian medical schools ....................... 20
  1.4.4 Types of simulation ................................................................................................. 20
  1.4.5 Simulation fidelity ................................................................................................... 21
  1.4.6 The Virtual Patient ................................................................................................. 21
  1.4.7 Virtual reality in medical education ...................................................................... 24
1.5 Utilising simulation for assessment of clinical competence .............................................. 26
  1.5.1 Validity of virtual technologies for assessment of competence ............................... 28
1.6 Virtual reality programme utilised for the current project ............................................. 30
  1.6.1 Description of the Otago Virtual Hospital (OVH) .................................................... 30
  1.6.2 Operating the OVH ............................................................................................... 31
  1.6.3 Current uses of the OVH ...................................................................................... 32
1.7 Aims of the Project ............................................................................................................ 36
Chapter Two: Methods

2.1 Participants

2.2 Testing

2.3 The clinical scenario

2.4 Analysis

  2.4.1 Quantitative scoring of performance

  2.4.2 Sub-score assessment and analysis

  2.4.3 Quantitative analysis of times taken

  2.4.4 Qualitative analysis

Chapter Three: Results

3.1 Quantitative analysis

  3.1.1 Demographic data of participants

  3.1.2 Overall scores achieved

  3.1.3 Achievement of complete diagnoses

  3.1.4 Achievement of safe effective management

  3.1.5 Time taken to make diagnoses

  3.1.6 Time taken to achieve correct and full diagnoses

  3.1.7 Time taken to complete clinical case

3.2 Qualitative analysis

  3.2.1 Thematic analysis of case transcripts

    3.2.1.1 Theme 1: Efficient ‘pattern recognition’ and illness scripts

    3.2.1.2 Theme 2: Elaborated knowledge

    3.2.1.3 Theme 3: Recognition of key features

  3.2.2 Thematic analysis of admission/discharge notes

    3.2.2.1 Overall structure of admission notes

    3.2.2.2 Structure and content of the ‘presenting problem’

    3.2.2.3 Diagnoses/Impressions

    3.2.2.4 Management plans

    3.2.2.5 Summary of differences between cohorts
Chapter Four: Discussion.................................................................91

4.1 Overview..................................................................................91
4.2 Evidence of validity.................................................................92
4.3 Validity of the scorecard as a tool for assessment of clinical performance........94
4.4 Specific assessment of clinical reasoning ability ................................................97
   4.4.1 Quantitative measurement of clinical reasoning .........................................97
   4.4.2 Measurement of efficiency by times taken....................................................97
   4.4.3 Assessment of clinical reasoning skills by analysis of transcripts.................98
4.5 Practicalities of implementation......................................................................100
   4.5.1 The value of sub-scores..............................................................................100
   4.5.2 The summary document: a useful insight into overall performance...101
4.6 Limitations to study..............................................................................102
4.7 Future directions.....................................................................................104
4.8 Conclusion..........................................................................................104

Bibliography.........................................................................................106

Appendices..........................................................................................112
List of Tables

Table 1. Demographic details of participants....................................................44
Table 2.1 Mean scores of clinical performance.................................................45
Table 2.2 Significance of differences between clinical performance scores......45
Table 3. Achievement of complete diagnoses by stage of training..................46
Table 4. Achievement of safe effective management by stage of training.......47
Table 5. Times taken to make diagnoses..........................................................48
Table 6. Times taken to complete clinical case...............................................49
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Miller’s Framework for Clinical Assessment</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2</td>
<td>OVH: Triage form</td>
<td>33</td>
</tr>
<tr>
<td>Figure 3</td>
<td>OVH: Taking a patient history</td>
<td>34</td>
</tr>
<tr>
<td>Figure 4</td>
<td>OVH: Obtaining virtual patient’s blood results</td>
<td>35</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Clinical Performance Scorecard</td>
<td>42</td>
</tr>
</tbody>
</table>
Definitions

Simulation

(n.) [sim’yələ’shan]

Etymology: Latin simulare, to imitate

- a method of representing the actions of one system by those of another, as a computer program that represents the actions of something in the real world. Simulation enables a computer to explore situations that might be too expensive, dangerous, or time consuming in real life.


Avatar

(n.) [ˈævə,tə:] 

Etymology: Sanskrit avatarati, he descends

1. a visible manifestation or embodiment of an abstract concept
2. a movable image that represents a person in a virtual reality environment


Fidelity

(n.) [fi’delitɪ]

Etymology: Latin fidēlis, faithful

- The degree of accuracy with which sound or images are recorded or reproduced.

Random House Webster’s College Dictionary

“Virtual reality is the first step in a grand adventure into the landscape of the imagination.”

Frank Biocca, Taeyong Kim, & Mark R. Levy.
Communication in the Age of Virtual Reality
Chapter One: Introduction

1.1 Overview

The competence of practising doctors has recently come under increasing scrutiny. With greater public awareness of the impact of medical errors, society now expects doctors to exhibit a sufficient and measurable level of clinical expertise. From a medical education perspective, providing valid assessment of competence is a complex task, and decades of research has been undertaken to ascertain optimal methods by which to achieve this. Now, emerging technologies such as computer-based simulations offer an alternative means of educating and evaluating medical professionals. It is possible that in the foreseeable future, evolving technologies such as virtual reality may revolutionise medical education and assessment.

The purpose of this study is to provide initial evidence for the validity of a virtual reality-based clinical case as a means of assessment of clinical competence. It also aims to investigate the capacity of the virtual case to evaluate clinical reasoning ability.

This introduction comprises a succinct review of all aspects pertaining to this project. The initial discussion reviews the definitions of clinical competence, with a particular focus on the theoretical basis of clinical reasoning. This is followed by an outline of how clinical competence may be assessed, including a review of the current methods of assessment of clinical reasoning ability.

An overview of simulation in medical education is then provided, emphasising the development of the ‘virtual patient’, as well as the incorporation of new technologies such as virtual reality into the realms of medical education. To conclude the literature review, a brief review of the current research into the validity of assessments using virtual technologies is presented.
The Otago Virtual Hospital (OVH) is the virtual-reality programme tested in this project. As an introduction to the study, the OVH is described in detail.

1.2 Clinical competence

Assessment of clinical competence is an essential component of medical education and is necessary for medical licensing by governing bodies. A widely accepted definition of professional competence, proposed by Epstein and Hundert in 2002 is ‘the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values and reflection in daily practice for the benefit of the individual and community being served.’ It follows that competence encompasses more than medical knowledge and skills, which are currently most often the basis for assessment.

In their seminal review of the topic, Epstein and Hundert defined six domains of clinical competence.²

The cognitive component of professional competence included not only core knowledge, but also the ability to continue to acquire and use information. It has been suggested that the cognitive processes of a competent physician not only rely on knowledge that is explicit (hence examinable), but also on less evident processes such as intuition, heuristics (mental shortcuts that allow rapid problem-solving, also known as rule-of-thumb strategies) and the ability of pattern recognition. These encompass many of the skills of clinical reasoning, discussed in detail later in the chapter. It follows that cognitive competence also includes the ability to use medical knowledge appropriately in real-life scenarios, and to be able to learn from experience.

Epstein and Hubert state that the capability to integrate is an essential component of professional competence. Integration is also considered a component of clinical reasoning ability. The competent clinician must be able to make sound decisions,
which requires the proficient connection of relevant information from various fields. Additionally, to achieve this in the clinical setting, a competent practitioner requires the ability to do this with incomplete information, to have the ability to manage uncertainty and ambiguity.\(^4\)

Clinical decision-making also has moral aspects. An essential element of clinical competence is the ability to show respect and care for patients, to ‘show willingness, patience and emotional awareness to use these skills judiciously and humanely.’\(^2\)

Other proposed domains of professional competence are technical competence (examination and procedural abilities), relationship skills (ability to communicate, to work with and teach others), and to perform suitably within the context in which health care is provided. This includes providing care that is appropriate to the clinical environment, and the judicious use of time appropriate to the setting.\(^2\)

Finally, on-going competence is determined by the ‘habits of mind’ of the professional; their ability to critically analyse their own performance, to recognise bias and error.\(^2\)

Clinical competence is thus a multi-faceted concept, and effective assessment should attempt to encompass all aspects.

1.2.1 Clinical reasoning

Clinical reasoning ability has been referred to as ‘the cornerstone of medical competence.’\(^5\) While many definitions have been proposed, it is generally accepted that clinical reasoning is ‘the cognitive process that underlies diagnosis and management of a patient’s presenting problem.’\(^6\)

Since clinical reasoning ability is considered an essential attribute of clinical competence, there has been over thirty years of research undertaken in an attempt
to accurately describe how it is performed. There is a vast body of literature in this area, so for the purposes of this project, only a brief review of relevant aspects will be discussed. This will include analytical theory (hypothetico-deductive reasoning), non-analytical methods – ‘pattern recognition’ (exemplar theory, illness scripts), as well as an overview of other theoretical aspects of clinical reasoning ability such as encapsulated knowledge, semantic transformation of information, and elaborated knowledge.

The original work on clinical reasoning suggested that diagnosis was the product of an analytical process; this was the basis of the hypothetico-deductive theory proposed by Elstein in 1978. Elstein theorised that doctors approach a clinical problem by generating preliminary hypotheses, and then proceed to collect information to either support or rebut the initial ideas, to deduce the correct diagnosis. However, subsequent research did not support this as a primary cognitive process crucial to clinical reasoning, since it appears that both novice and expert clinicians can use this process identically; the key difference being that the expert can generate more accurate diagnoses initially. It would seem that the experienced doctor was able to unconsciously identify more likely diagnoses before analysing all available information, and this led to the idea of non-analytical clinical reasoning.

Suggested to be the ‘central component of diagnostic expertise at all levels’, non-analytical clinical reasoning theory was proposed partially to explain the obvious ability of the expert to arrive at more accurate diagnoses more quickly than a novice clinician. If it could be assumed that experienced physicians would have a broader knowledge base from which to test hypotheses, a purely analytical process would take longer, as they would have more information to access. In fact, what was observed that in most cases, the expert clinician was able to reach diagnoses more quickly than inexperienced colleagues, and it was also shown that accurate diagnosis and speed were positively correlated. The proposed explanation for this was that of ‘pattern recognition’, an idea that has since been further developed in clinical reasoning research.
When a skilled doctor is confronted with a routine presentation, the cognitive processes involved often seem to be tacit - the doctor ‘just knows’ the likely diagnosis, without being able to make explicit their underlying reasoning. An everyday example of this process is the ability to identify a dog from a cat. Almost everyone can do this, but explanation of how this is achieved becomes quite complex, (Professor Tim Wilkinson, personal communication, November 2012). From this observation arose the ‘exemplar theory’ and ‘illness scripts’. In essence, experience with numerous previous examples (or exemplars) enables the expert to subconsciously match and identify a presentation, without needing to intentionally analyse individual aspects to make a decision.

The concept of ‘illness scripts’ was coined by Feltovich and Barrows, and was adapted from the script theory used in cognitive psychology. An illness script is a schema of knowledge about a disease process, including ‘what is normal and what acceptable variations are’, and is a result of accumulated prior experiences of the illness. When an experienced physician identifies salient features in a clinical presentation, applicable illness scripts are activated. The doctor can then assess how well the patient presentation matches known illness scripts, and this assessment of fit is believed to usually occur in a subconscious, non-analytic manner. Deliberate analytical reasoning is only initiated if there are multiple scripts to compare, or with a vague undifferentiated presenting complaint.

It is proposed that illness scripts form as a result of encapsulation of knowledge. Encapsulated information is information that is stored but not routinely accessed. It involves the ‘chunking’ of the basic underlying details into higher-level concepts or models. An example of medical knowledge encapsulation, (as illustrated by Schmidt and Rikers, 2007) is the expert clinician’s use of the term ‘sepsis’, to simply explain all of the clinical signs and symptoms, as well as the pathophysiological basis of systemic infection.

Schmidt and Rikers proposed that formation of illness scripts was part of four stages of development of medical expertise. They suggest that initially, understanding of
disease is primarily pathophysiological. This knowledge then becomes encapsulated into ‘diagnostic labels, syndromes or high level models’ as described above, then experience and context allows the linking together of these concepts into illness scripts. The illness scripts are then further refined by experience and memory of multiple exemplars.\textsuperscript{12}

More recent literature suggests that the non-analytical methods such as illness scripts are likely to be only part of a ‘highly complex and multifaceted process.’\textsuperscript{9} It is now thought that clinical reasoning involves multiple simultaneous cognitive processes that occur in a non-linear manner. Non-analytical methods are a significant component, but other cognitive processes seem to occur in more complex or ambiguous cases. It has been shown that when other processes fail, expert doctors may resort to physiological mechanistic based reasoning, or utilise other forms of encapsulated knowledge.\textsuperscript{9} Encapsulated knowledge is also considered valuable when assessing the fit of illness scripts, as the underlying scientific principles constrain what is physiologically feasible for a given problem.\textsuperscript{11}

This resulted in a conceptual change in clinical reasoning theory, from that of it being a process that could be learned (or taught), to it being the desired consequence of experiential learning. To become skilled in clinical reasoning, a practitioner needed to access organised knowledge based on multiple previously encountered exemplars, which is obviously dependent on clinical experience.\textsuperscript{3}

However, clinical experience does not inevitably result in skilled diagnostic thinking; for this to occur is thought to require ‘elaborated learning’. Elaborated learning has been proposed to be the ‘foundation for problem solving and probably effective clinical reasoning.’\textsuperscript{13} It has been observed that the expert doctor is better able to activate their illness scripts than a novice colleague. The experienced clinician can quickly infer the clinically important features, and then produce and compare more accurate hypotheses.\textsuperscript{11} This ability to ‘elaborate knowledge’ is the capability to retrieve and apply knowledge, as it is relevant to different clinical situations.
Elaborated learning then describes the ability to integrate information, deemed an attribute of clinical competence.\textsuperscript{13}

A variation on this concept, ‘elaborated knowledge’, was described by Bordage in 1994.\textsuperscript{14} He proposed four models of knowledge arrangement; these were reduced, dispersed, elaborated and compiled knowledge. Reduced and dispersed knowledge structures are thought to be associated with poor diagnostic ability, while elaborated and compiled knowledge are considered requisite for expert diagnosis.\textsuperscript{14}

To have elaborated knowledge first requires the capability to describe clinical findings as abstract qualitative terms (semantic qualifiers). For instance, symptoms appearing over several months can be defined as ‘gradual onset’ and pathological signs evident on both sides of the body as ‘bilateral’. This ability to interpret and transform the presenting problems into key abstract concepts is known as the semantic transformation of information.\textsuperscript{9} The relationships between the semantic qualifiers can then be used to assess possible diagnoses. The expert with elaborated knowledge does not consider symptoms and signs in isolation, but instead compares associated presenting issues to their own underlying schemas of organised ‘elaborated’ knowledge structure. Elaborated knowledge involves connecting the patient’s findings with underlying knowledge, recognising not only the cause, but the relationship between symptoms.\textsuperscript{14} An example of this would be the diagnosis of the patient who presents complaining of experiencing shortness of breath on a number of occasions, who also notes an unusual awareness of their heart beating at times. The experienced clinician converts the presenting information into the semantic qualifiers; ‘episodic’, ‘dyspnoea’, ‘palpitations’, and for this combination may quickly come up with a possible diagnosis of paroxysmal atrial fibrillation (AF).

However, if the student or clinician exhibits reduced knowledge, this results in the inability to connect the issues of the patient with their own existing knowledge, (i.e the structure of knowledge is reduced, unable to be accessed). Dispersed knowledge occurs when the doctor is able to retrieve ample information, but is not able to evaluate it in a united manner. Dispersed knowledge often leads to multiple
diagnoses for each presenting symptom, but no unifying diagnosis. Compiled knowledge is encapsulated knowledge, which allows the highly experienced clinician rapid tacit access to underlying information without the need for elaboration.\textsuperscript{14} Using the above example of atrial fibrillation, the student with reduced knowledge would be unlikely to suggest a diagnosis at all, and the student with dispersed knowledge may suggest diagnoses generated from an isolated symptom, for example ‘chronic obstructive pulmonary disease’ or ‘Guillian-Barre syndrome’. These diseases do indeed produce dyspnoea, but are diagnoses that do not fit well with other presenting information. The clinician with compiled knowledge makes the diagnosis of paroxysmal AF without any conscious elaboration of the underlying thought processes.

In summary, current understanding of clinical reasoning suggests that it involves utilisation of multiple cognitive processes, and requires access to knowledge stored and organised in different forms. Individual pieces of medical knowledge become related into semantic networks, which are in turn encapsulated into illness scripts. Multiple exposures to clinical examples enable rapid recognition and diagnosis by accessing these scripts. To do this, the expert clinician must first convert presenting information from the patient into meaningful clinical concepts (the semantic qualifiers), and then assess the relationships between these, and compare them to their ‘stored mental representations of disease entities’.\textsuperscript{15} Clinical reasoning is necessarily so complex because, as stated by Elstein in 2009 (over 30 years after his initial work on hypothetico-deductive theory), ‘intuition is not perfect and that rational thought is too time-consuming.’\textsuperscript{16}

\textbf{1.3 Assessment of clinical competence}

Essential for the assessment of clinical competence is the development of valid, reliable tools with which to evaluate the performance of students and doctors.\textsuperscript{17–20} A framework for assessment of competence, clinical skills and performance was
proposed by Miller in 1990, who devised what is now widely known as ‘Miller’s pyramid’, (figure 1).¹⁸

Figure 1. Miller’s framework for clinical assessment.

The substantial base of the pyramid is compiled of the ‘knows’ and ‘knows how’ layers: this contains the knowledge base of the practitioner, and also their ability to use and apply the knowledge as required to solve a clinical problem. Lying above these layers is the ‘shows how’ layer, reflecting the actual performance of the examinee with ‘patients’ when observed. Above performance sits action, and the ‘does’ layer is situated at the peak of the triangle, where ideally assessment of the doctor’s independent clinical behaviour could occur. As stated by Wass et al (2001) in the Lancet; “Assessment at the apex of Miller’s pyramid, the “does”, is the international challenge of the century for all involved in clinical competence-testing.”²⁰ Miller himself contends that the goal of assessment of clinical competence should be to test in the upper levels, although also states that “no single assessment method can provide all the data required for anything so complex as the delivery of professional services by a successful physician.”¹⁸
The development of tools to assess clinical competence must take into account the fundamental aspects of assessment: validity, reliability, blueprinting and standard-setting.\textsuperscript{20}

A valid test is one in which the examination does indeed assess the aspect that was intended to be tested.\textsuperscript{19} Different aspects of validity can be assessed; these include face validity, content validity, predictive validity, construct validity and concurrent validity.\textsuperscript{21} Face validity simply indicates that a test appears (at face-value) to be testing what was intended, and a test with content validity is able to evaluate the expected content. Predictive validity signifies the ability of a test to predict future performance, and construct validity represents the ability of the test to evaluate an abstract construct, such as clinical reasoning ability. Concurrent validity indicates that a test compares acceptably with another established evaluation.\textsuperscript{21} Ideally, an assessment tool should encompass many of these forms of validity.

Validity can be inferred by a variety of measures. Most simply, if the test is a valid test of competence, it would be expected that experts will out-perform novices. Within the test, participants should score highly on items if they have mastery of that particular subject. Another means of ascertaining validity is by comparing tests that are presumed to evaluate the same attribute; valid tests will result in a high correlation of scoring.\textsuperscript{19}

Reliability of an assessment tool describes the reproducibility of the scores produced. While influenced by many issues, key problems for reliability are mainly due to differences resulting from variations between examiners, and the issue of case specificity. In essence, increasing the number of examiners for assessment improves reliability, and testing across many cases is necessary because knowledge and hence competence is case-specific.\textsuperscript{20}
Blueprinting describes the process of developing the assessment to reflect the learning objectives, and standard-setting involves clear outline of the expected endpoint, i.e. the standard to be attained at which competence is achieved.\(^\text{20}\)

Most medical schools currently use a mixture of assessment modes.\(^\text{17}\) Written assessments may include multiple-choice questions (MCQs), short answer or essay type questions, or the more recently developed key-feature or script-concordance questions.\(^\text{22,23}\) These are further discussed in the following section. Other forms of assessment include evaluation by supervising clinicians in the practice environment. This may be in the form of a global appraisal, or as a result of planned observation in the clinical setting, such as the ‘long case’, sometimes with an accompanying oral viva voce, or the ‘mini clinical evaluation exercise (mini-CEX).\(^\text{24}\) Various forms of clinical simulation are often also utilised, the most common being the objective structured clinical examinations (OSCE), which uses actors who play standardised patients. Assessment by types of high fidelity simulation (generally mannequins or other multimedia type environment) is progressively being used in the life support, anaesthesia, intensive care and surgical fields.\(^\text{17}\)

In a recent review of current modes of assessment of medical competence, Schuwirth (2004) outlined the most consistent findings of research in the area.\(^\text{19}\) Once again, the most significant issue appears to be that of case-specificity, that demonstrating knowledge or performance in one area does not predict a student’s ability in another. Also of note, is what is described as the ‘idiosyncrasy of problem-solving.’ It appears that there is rarely a consensus among experts about the correct process by which to solve a given medical problem, so evaluating competence by assessing process rather than outcomes lacks validity. Schuwirth also argues that the format of an assessment is less important than its content, and that global impressions (if given for directly observed performance, with multiple examiners) were as reliable as checklists for scoring.\(^\text{19}\)

The purpose of assessment is not only to protect patient safety, but also drives student learning.\(^\text{19}\) Results of assessment can supply the examinee with information
about their deficiencies, enabling them to continually improve. Assessment also provides information that can be used for curriculum development.\(^2\) It is then obvious that assessment of clinical competence is a critical aspect of medical education, however current means of evaluation are far from comprehensive.

1.3.1 Specific assessment of clinical reasoning ability

Assessment of clinical reasoning ability is considered an essential aspect of medical education,\(^{15,25}\) and this is emphasised in many major medical school curricula worldwide. Given that there has been many decades of debate about the nature of clinical reasoning, developing an effective tool with which to measure it has been even more problematic.\(^{26}\) Currently, there is not a universally accepted approach to assessment of clinical reasoning that is considered to be valid and reliable.\(^{25}\)

Most current modes of assessment are considered to provide poor tests of clinical reasoning ability. It is acknowledged that MCQs are problematic with respect to assessment of diagnostic ability. Subjects may recognise the correct option for a given question yet be unable to generate it independently, ('cueing').\(^{17}\) Other forms of written testing such as short answer or essay type questions may be appropriately designed to test clinical reasoning, but are obviously time-consuming, and are dependent on reliable grading.\(^{17}\)

Key feature problems, introduced by Bordage and Page in 1987, were developed for the qualifying exam in medicine for the Medical Council of Canada as an alternative means of assessing clinical decision-making skills.\(^{22}\) Decision-making reflects a crucial part of the process of solving clinical problems. As outlined above in the review of clinical reasoning, the effective practitioner must first identify the clinical problems from the presenting information. To then resolve the patient’s issues, they then must supply potential diagnoses and solutions, deciding on optimal management. Clinical decision-making thus reflects the outcomes of this process.\(^{27}\) Since key feature problems were designed primarily to assess clinical decision-making, they
reflect the application and integration of information, as well as the candidate’s judgement.\textsuperscript{27}

Key feature problems are generally offered in a written (short answer) or computerised format. A clinical scenario is provided and then questions posed that focus only on identification of relevant ‘key features’ – crucial elements required for diagnosis and management. These may be either clinical decisions, for instance ‘\textit{include pulmonary embolism in the differential diagnosis}’, or clinical actions such as ‘\textit{order CTPA}’ (computed tomography pulmonary angiogram).\textsuperscript{28} A key feature problem may require the subject to either elicit or interpret data, depending on which aspect is considered to essential to the given scenario. Since the questions focus only on tasks considered critical to resolve the particular clinical case, the key features are inherently specific to the clinical problem. However the short structure allows sampling over many cases, enhancing content validity. Different scoring systems have been developed, most often with simple weighting schemes and sometimes incorporating penalties for dangerous decisions.\textsuperscript{28} Key feature questions require consensus from expert authors on what constitutes the critical steps for the specific problem.\textsuperscript{22,28}

A recent evaluation of the efficacy of key features problems indicated that they are likely to be ‘\textit{more cognitively complex}’ compared with MCQs, which implies they better assess skills like clinical reasoning, a ‘\textit{higher level cognitive process}’.\textsuperscript{27} Advocates of key features problems assert that are unique in that they enable the evaluation of the student’s ability to ‘\textit{apply knowledge and training to real world problems in an integrated and appropriate manner}’.\textsuperscript{27}

One of the more recent developments in clinical reasoning assessment is the script concordance test.\textsuperscript{23} Originating from illness script theory, it is an assessment method developed in the 1990s to test the ability to interpret clinical data in the context of a poorly differentiated case scenario. In written form, a brief clinical problem is outlined. It is followed by a selection of proposed diagnoses, each accompanied by a further piece of clinical datum. Subjects are required to assess the effect of the
additional material on the probability of the associated diagnosis being accurate. Scores are generated by comparison to expert opinion.\textsuperscript{23} The script concordance test was specifically designed to measure the effect of experiential knowledge on diagnostic ability, and is in line with current theory on the nature of the development of clinical reasoning. There is emerging support for its validity, however it is not yet widely used for summative assessment.\textsuperscript{23}

Clinician-observed forms of assessment such as the long case have traditionally been the mainstay of assessment of clinical thinking.\textsuperscript{29} The long case is usually an unstandardized examination, with a single or small number of cases, and often a single examiner. Problems arise with bias, subjectivity and case specificity, and while still used in the absence of better alternatives, the long case is considered by many to be unreliable means of assessment.\textsuperscript{26,29}

Observed clinical simulations, primarily in the form of OSCEs have become increasingly popular, and are proven to be an efficient means of assessing history taking and clinical examination skills, but also are still somewhat limited in their capacity to evaluate clinical reasoning ability.\textsuperscript{26} Other forms of more elaborate simulation designed specifically to assess clinical reasoning are still in early stages of development.

Further work is then needed to develop an effective, valid and reliable tool to evaluate clinical reasoning ability. Assessment in medical education is becoming progressively less dependent on written examinations. There is a trend towards performance-based appraisals, as there is a general consensus that written tests are a poor measure of application of knowledge and clinical ability.\textsuperscript{25} Performance based assessment, including utilising standardised patients and simulations, are assumed to provide a better means of evaluation of clinical reasoning. However, this has not yet been subjected to adequate critical scrutiny.\textsuperscript{25}

As discussed, the most recent work on clinical reasoning indicates that it is an ability that is closely associated with the way in which knowledge is stored and accessed. It
is currently thought that new assessment methods need to be developed that have the capacity to examine both the underlying knowledge and its application in the clinical setting. It has been suggested that clinical reasoning, then, should not be assessed alone, but must be combined with evaluation of medical knowledge.\textsuperscript{15} It becomes then inherently difficult to evaluate the construct validity of tests of clinical reasoning ability, as outcomes may reflect knowledge alone, rather than application.\textsuperscript{25}

Since clinical reasoning ability is principally dependent on the capacity to organise and retrieve knowledge, this naturally must be specific to information content. A doctor who appears competent in a given area may then demonstrate less skill in clinical reasoning when presented with a different clinical problem. To avoid this problem of content specificity, it follows that it would be essential that any new form of assessment contain multiple cases.

In conclusion, while extensive research has been undertaken to develop and test new tools to assess clinical reasoning, there is as yet no well-validated and universally accepted instrument available.
1.4 Simulation in medical education

The use of simulation for training purposes is well established in many areas, particularly in high-risk professions such as aviation, space exploration, warfare and the nuclear power industry. Over a decade ago, it was predicted that ‘the future of medical education is no longer blood and guts, it is bits and bytes.’

Nowadays, for the purposes of medical education, modern educational technologies can provide computer-based tools that can reproduce, for example, realistic physiological and anatomical models on which students may learn and practise new skills safely. Rapid advances in technology now enable higher quality and complexity of simulation to be developed. Consequently, simulation now provides a safe and educational environment in which students can be challenged with realistic experiential learning.

1.4.1 Historical perspective

When medical schools worldwide began to include simulation into their programs, the evidence base for the effectiveness of their use was scanty. Much of the existing research on its use in medicine was criticised for poor methodology, as well as inadequate alignment with accepted educational theory. However, more rigorous research already supported simulation based training in other fields, such as aviation and military.

Published reviews of the literature now suggest that simulation is having an increasing impact on medical education, and the technology is generally well accepted and liked by medical trainees. This is thought to be because the format of simulation is usually more realistic than other forms of education; it offers ‘an enhanced environment for experiential learning and reflective thought.’ While it has been suggested that there is not yet sufficient research of adequate quality available to unequivocally support its effectiveness as an educational tool for all
aspects of medical education, simulation is a rapidly evolving field and there is increasing evidence for its value. A 2011 meta-analysis that compared traditional clinical teaching with simulation-based methods for skills acquisition (e.g. advanced cardiac life support, thoracocentesis, cardiac auscultation) showed that simulation produced indisputably superior outcomes. Another 2011 meta-analysis demonstrated that ‘technology-associated simulation training,’ was associated with a large effect on medical skills, knowledge and behaviours as well as a moderate effect on patient care. With these conclusions, the authors then questioned the need for any further work to demonstrate the value of simulation training, (when compared to no other interventions).

However the use of simulation for assessment purposes has still not been consistently well validated. For medical educators, more rigorous studies are required to support the introduction of these technologies to provide the most effective learning and assessment environments. In many centres, simulation-based education programmes are rapidly being developed, and included into clinical education curricula. Further work is needed to establish how simulation should best be incorporated into existing programmes to enhance clinical learning.

1.4.2 Factors driving development of simulation

Initial resistance to the development of simulation was perhaps due in part to the archaic impression that it was merely ‘gaming’, and that simulation technology was not appropriate for serious educational use. Refuting this view was Graafland et al, 2012, stating ‘to date, many medical professionals still have a rather out-dated view of the average ‘gamer’, as being someone who is too young to vote, afraid of daylight and killing mystical dwarfs in games like World of Warcraft® (Blizzard Entertainment, Versailles, France) in their parents’ basement. Contrary to this view, adults are avid users of digital devices.'
It is acknowledged that the use of educational technologies such as simulation provides some obvious advantages for the medical learner. As well as allowing authentic yet safe learning environments that can be customised, these tools can also enable recording of the learner’s behaviour and results. There may be less dependence on available expert tuition or clinical cases, and also, often these tools can be used at any time or place, any number of times. In the hospital setting, patient care must be prioritised over student learning. Consequently, medical students rarely have the opportunity to integrate their knowledge by taking responsibility for clinical decision-making, especially in the management of critically ill patients.

Simulation has been promoted as an ideal environment in which students can gain experience in managing complicated or critical clinical situations without risk of harm to patients. The ethics of medical practice deem that doctors must ‘first do no harm’; the protection of patients is paramount. Practitioners are ethically bound to provide optimal care, but this must be balanced with the requirements of training.

The apprenticeship model that requires students to develop their skills by practising (under supervision) on real patients is now less acceptable to the public. Over half a decade ago, Ogden et al stated ‘in the near future it should become unthinkable for trainees to practice and acquire basic skills on patients,’ a sentiment expressed also in emergency medicine literature; that the ‘see one, do one, teach one’ doctrine should have been abandoned long ago.

Vocational specialties such as anaesthesia, and emergency medicine, have readily embraced simulation for training purposes, as these specialities require time-critical clinical decision making, for which practice using simulation offers evident benefits. It has been proposed that simulation may also be valuable for the training of teams of health care professionals to function optimally in crisis situations. Internal medicine has also begun to adopt simulation technologies. While usually utilised to teach invasive procedural skills such as cardiac stent placement, there is increasing recognition of the value of simulation to teach clinical reasoning and communication skills relevant to internal medicine.
The drive to develop simulation-based learning and assessment for medicine is compelling, as the competence of new graduates is increasingly coming under question. The changing structure of the medical workplace, with decreasing work hours, results in less clinical experience. Junior doctors are often responsible for complex and critical patients, with senior supervision not always immediately available. With larger medical student numbers, and concurrently a greater emphasis on patient safety, ‘practice’ is not possible, so trainees may be unable to learn how to make decisions in critical situations.47

A recent disturbing analysis used clinical simulation to assess the abilities of medical students, residents and experts recruited from trauma and surgical critical care rotations to manage basic clinical scenarios relating to critical care. This encompassed testing of 140 cases, with cases such as atrial fibrillation, acute myocardial infarction and sepsis. Results from this study indicated that compared with experts, participants from all other levels of training showed a significantly poorer performance. The authors concluded that this was not due to a flaw in the assessment, but that ‘we are not teaching medical students and junior doctors a cognitive framework to function in these situations...’47

Another advantage of this developing technology is that both learning and assessment may be individually adapted for specific students, or alternatively, may be standardized as desired.32 For these reasons, the use of simulation for competence assessment is gaining support. ‘This educational method will become the standard for competency assessment in the near future and be as routine for the continuum of physician education as it is for aviation and other high stakes professions.’41 However, as discussed in detail below, the validation of simulation-based assessments is still requires further research, and is sometimes opposed by the physicians responsible for evaluation of trainees.48
1.4.3 The local perspective: Simulation in Australasian medical schools

Due to associated expenses, most simulation technologies have been developed in North America and Europe. However, more recently, Health Workforce Australia has been attempting to develop an organized approach to simulation in Australian medical schools, in order to coordinate and improve existing simulation programmes.  

1.4.4 Types of simulation

While there are multiple kinds of educational technologies available, the Association of American Medical Colleges considers that there are three predominant categories of applications used in medical education. These consist of:


   Computer aided instruction (CAI) is defined as information provision using a computer as a medium. Typically these are relatively inexpensive programmes, and can be useful to provide visual representations of complicated processes. CAI often allows the student to interact in a limited way with the program in order to learn information, but not in a decision-making capacity.

2. Virtual patients (VP)

   This is a more specialised technology, in which a simulated patient is provided, and the student is required to act as the medical professional to manage a clinical scenario. This delivers a more realistic clinical experience, although without any physical interaction being required. Advantages of the VP include the opportunity for the student to experience the various elements of patient management, such as history taking and ordering investigations, and more advanced systems will allow the VP to respond to the student in real-time, and react to interventions. The VP is easily adapted to provide alternate scenarios, levels of difficulty or learning objectives. VP
programmes often have constraints on the level of fidelity possible, and may be expensive to develop initially.

3. Human patient simulation (HPS)

Generally this involves the use of mannequins, or models of a body part, and is often for the purposes of procedural training, or to replicate a clinical environment. This form of simulation offers the best level of fidelity, and requires the learner to be physically involved in the experience. HPS necessitates a high level of cost and expertise, and because of manufacturing limitations, a smaller range of medical scenarios may be utilised. There are significant on-going costs associated with the use of HPS, as this type of simulation requires continuing financial and human contribution.\(^32\)

1.4.5 Simulation fidelity

Simulation fidelity can be thought of the authenticity provided by the simulation experience. The relationship between level of fidelity and learning outcomes is not necessarily a simple correlation, as skill improvements have been produced with low fidelity simulation,\(^32\) and it has been suggested may be that high fidelity systems may more beneficial for the expert learner, while the novice can still learn effectively from the less expensive and complex forms of simulation.\(^32\) Hand-held gaming consoles (such as the Playstation Portable and iPod Touch) are commonplace and relatively inexpensive, and offer high-fidelity experiences for entertainment purposes. This may possibly generate expectations of higher fidelity in simulations than is currently feasible.\(^48\)

1.4.6 The Virtual Patient

Although almost all types of simulation are now being utilised in medical education, the focus of this review is primarily on the use of the virtual patient (VP). VPs can be defined as a ‘specific type of computer program that simulates real-life clinical
scenarios; learners emulate the roles of health care providers to obtain a history, conduct a physical exam, and make diagnostic and therapeutic decisions.’

The range of what may be described as a VP in the literature is quite broad, but generally excludes advanced forms of simulation such high-fidelity mannequins, part-task trainers, and specialised haptic surgical and procedural trainers. This is because the key function of the virtual patient is that that it is able to respond to user input, as opposed to other kinds of computer based learning such as the CAI already outlined. Using a VP, the clinical scenario develops as the user interacts with the programme, and the user is required to make diagnostic and management decisions.

In various forms, the VP has been in existence since the 1970s, but has been slow to be incorporated into medical curricula. As recently as 2007, only 26 out of 108 surveyed medical schools in the United States and Canada were in the process of developing virtual patients. Technological limitations to construction have curbed development of the VP, as well the large expense and duration associated with their manufacture. Primarily, while the cost of development has been prohibitive, and there was also thought to be insufficient support for their educational efficacy. A 2010 meta-analysis of the literature on VPs demonstrated overall a large positive learning effect of VPs compared to no intervention, but a small effect compared with other types of teaching. This meta-analysis also illustrated that the majority of the published research was not able to be included due to limitations of study design.

However, evidence for the efficacy of VPs in medical education has been slowly accumulating. In a recent (2010) randomised controlled trial, Botezatu et al demonstrated that VP-based learning and assessment produced superior retention of knowledge in cardiology and haematology modules, compared with conventional teaching methods. This trial, with eleven others, was included in the most recent meta-analysis of the efficacy of VPs in medical education that was published in 2012. The analysis included only randomised studies, and concluded a clear positive overall
effect favouring VPs. Sub-analysis indicated the greatest effect on clinical reasoning ability.\textsuperscript{54}

Of all the emerging simulation technologies, it would seem that VPs have some distinct advantages. As stated, changes in the way medicine is practised has resulted in less patients being available as inpatients for student learning.\textsuperscript{50} It has become difficult to provide learners with the opportunities for them to develop skills in an experiential manner. The concurrent demand in medical education for standardised examinations means rigorous assessment of medical student abilities is challenging, especially when evaluating the management of critical situations.\textsuperscript{39} The VP is ideally suited to enhance clinical learning, by presenting an alternative setting for acquisition and evaluation of clinical skills and decision-making. Another obvious benefit of the VP is that it may be designed to present clinical scenarios as needed – common, rare or difficult presentations may be developed as required.

Medical students have been reported to find virtual patient simulation valuable, accepting the use of the novel technology for both education and assessment purposes.\textsuperscript{55}

In a 2009 review by Cook et al, it was proposed that the predominant educational advantage offered by VPs is to facilitate the development of clinical reasoning.\textsuperscript{49} As described in detail above, expert clinical reasoning is a complex process. It involves the ability to utilise existing knowledge, to gather and collate relevant information to correctly diagnose and manage a clinical case. The clinical reasoning literature suggests that this ability develops through practice, and by exposure to multiple different clinical examples. VPs could be well utilised to facilitate this.\textsuperscript{49} In addition, open-ended VPs allow the recording of the pathway taken to resolve the clinical problem, offering a unique insight to the nature of clinical reasoning.\textsuperscript{39} It then follows that VP-based assessment then requires development of outcomes that can measure clinical reasoning ability.
Current opinion suggests that VPs are already the ideal replacement for paper cases, but then need to be developed with specific goals in mind, whether that be to facilitate or assess clinical reasoning ability, to allow repetitive practice of key clinical skills, or to offer a standardised experience of unusual or difficult presentations.

Of note, most forms of VP still involve using some form of text-based interface to take the patient history, although attempts have been made to utilise voice recognition technology. A recent interesting development has been the use of the combination of a human standardised patient with a form of high fidelity simulation (either part-task trainers or virtual reality systems) – this is known as ‘patient-focused simulation.’ Proponents of this approach suggest that it adds the authenticity of a real human encounter, and maintains an essential focus on communication and patient-centred care.

1.4.7 Virtual reality in medical education

When first introduced, virtual reality was defined as a technology that could ‘present virtual objects or complete scenes to all human senses in a way identical to their natural counterpart.’ Potential applications of this in medicine included the reconstruction of radiological images for diagnostic purposes, pre-operative planning, and rehabilitation systems for patients, as well as being beneficial for education purposes. The modern interpretation of the term ‘virtual reality’ is that of the experience of operating within a virtual world: an online computer-generated multimedia environment where users can be ‘present’ and interact with others via an avatar, a visual depiction of themselves.

‘Virtual reality’ is currently used to describe an increasingly wide array of computer-generated or mediated environments, experiences and activities ranging from the near ubiquity of video games, to emerging technologies such as teleimmersion, to technologies still only dreamed of in science fiction and only
encountered in the novels of William Gibson or Orson Scott Card, on the Holodeck of television's Star Trek, or at the movies in The Matrix of the Wachowski brothers, where existing VR technologies make possible a narrative about imagined VR technologies. The term ‘virtual reality’ covers all of this vast, and still rapidly expanding, terrain. Derek Stanovsky.\textsuperscript{60}

There is often an association of virtual reality with gaming for entertainment purposes. Interestingly, a recent systematic review indicated that ‘serious games’ (games that resulted in useful skills or knowledge applicable to medical practice) could be of educational benefit, particularly for surgical training.\textsuperscript{38}

The most widely available and well-known virtual reality programme is Second Life. Second Life is a computer programme developed by Linden Labs Inc. in 2003 (\url{http://secondlife.com/}). It is a web-based virtual world that can be accessed by users from any location.

While it was over a decade ago when it was recognised that virtual reality based simulations could be of benefit for training emergency medicine doctors,\textsuperscript{42} it has only been fairly recently been utilised for other medical educational purposes.\textsuperscript{59} There is still very limited published research available evaluating its effectiveness. Recent applications included the use of a Second Life-based simulation to apply a mock examination to a small number of Emergency Medicine residents. Although none of the participants had prior experience with the virtual reality programme, the majority found it easy to navigate within the programme. Most importantly, it was perceived as more realistic than the standard mock oral examinations, and participants believed that it should be utilised in other areas of medical education. All participants felt it was a fair and objective means of assessment.\textsuperscript{61} From this small study, the indications would be that utilising Second Life as a modality for virtual patients would be well accepted by trainees. This view has been supported by recent work from the University of Sydney where development of a virtual hospital for medical education purposes has been well received by participants.\textsuperscript{62}
Second Life has also been utilised as a means by which to deliver a postgraduate continuing medical education programme. Again, one of the key findings was that the modality was well liked by the participants, all stating that the virtual reality based simulation was better than other internet-based options, and the majority stated that it was ‘as good as, if not better than’ traditional formats of delivery.63

While as yet in the earliest stages of development, Second Life and other immersive 3D experiences are posed to offer innovative forms of medical education in the near future.64 It was recently predicted that ‘the ‘Millennial Generation’ of students will no longer be served by the passive approaches and technologies of the past.’65 It is evident that development of these new educational tools will require a collaborative effort between contributing physicians, medical educationalists and the virtual reality system developers.66

1.5 Utilising simulation for assessment of clinical competence

In medicine, efficient yet thorough assessment is essential, in order to protect patient safety. Testing cannot merely cover knowledge of medical sciences, as the practice of medicine requires the ability to integrate and apply information, to use clinical reasoning and make management decisions. Currently used assessment tools lack the capacity to test these competencies in an objective and reliable manner.17,19–21

Arguably, the most important component of a medical assessment tool is its predictive validity. This is a key aspect in effective robust medical assessment, as testing needs to not only measure the current abilities of the subject, but also how they are likely to function in future situations.21 As simulations most closely replicate real practice compared to other forms of assessment, it is hoped that they may then provide the best predictive validity.
Various forms of clinical simulation have been utilised for assessment in medicine. Written versions of simulations have been part of medical assessment for some time. Early versions of these generally encompassed a written clinical scenario, requiring the student to then proceed through the case, eventually forming a diagnosis and management plan. Scoring of their performance was generally based on a predefined algorithm, with points awarded for each decision made. Representative of these written simulations were the Patient Management Problems (PMPs), prevalent in the 1960-70s. These have since largely been abandoned, as further research has indicated they have poor validity, in part because by the nature of their design, they award thoroughness and penalise efficiency, which is often the hallmark of an expert.67

More recent developments include the Key Features Problems,22,27,28 discussed previously. These can be considered a type of written simulation, although they lack some of the authenticity of a true simulation, as they only focus on selected critical features of a case. While the Key Features Problems are thought to have good construct validity, they are still only able to offer a linear simulation, i.e. there is usually often only one pathway through the case.67

Directly observed assessments such as OSCEs may also be considered as clinical simulations, as they generally involve the use of standardised patients (a trained actor playing a patient). An advantage of OSCEs is that each station is structured to be of short duration, so it is usual to increase the sampling with multiple stations, hence decreasing the problem of case specificity. OSCE stations are usually designed to test one specific skill at a time, so do not test integration well, and have been shown that in most cases, still test a large component of the knowledge, rather than its application.67

However, all forms of clinical simulation-based assessment are subject to case specificity. In a description of clinical simulation based assessment, Schuwirth argues that the ideal is to achieve a balance between authenticity (important for face validity), and brevity (to allow greater sampling of cases.)67 However, he also states
that ‘no single simulation-based method will capture the full scope of medical competence’.\textsuperscript{67}

It has been suggested that clinical simulation is the ideal instrument with which to test in the upper parts of Miller’s pyramid; a previously discussed framework for clinical assessment.\textsuperscript{18,67} Almost all forms of current written simulation assessments lie in the ‘knows’ and ‘knows how’ base of the triangle, while OSCEs and other observation based assessments attempt to evaluate the student’s ability in the ‘shows how’ section.\textsuperscript{67} It would be most desirable to assess within the peak of the triangle, (where the student ‘does’). This would likely provide the best predictive validity, an insight to the student’s future independent clinical behaviour.\textsuperscript{18}

Current assessments cannot easily evaluate the performance of a student in a real clinical setting, due to ethical constraints for patient safety. However, as predicted by Miller over two decades ago, more advanced simulations that can include dynamic aspects such as time, and better approach ‘reality’ may offer a valid alternative in the future.\textsuperscript{18} It would seem then, that simulations utilising newer technologies may be the best-placed to offer more authentic and comprehensive forms of assessment.\textsuperscript{68}

\textbf{1.5.1 Validity of virtual technologies for assessment of clinical competence}

To achieve this, the virtual patient (VP) has also been used not only for teaching, but also as an assessment tool. Part of the perceived advantage offered by VP based assessment is thought to be in its ability to test abstract concepts such as clinical reasoning ability. The nature of VP assessment means that the ability to integrate information and make clinical decisions can be evaluated in a more realistic fashion than current available medical assessment tools.\textsuperscript{21} While there are increasing efforts to provide evidence to support the use of technology based assessments,\textsuperscript{68} there has been very little published literature specifically evaluating the construct validity of those using the VP-based format.
As a form of assessment, VPs have good face validity (i.e. they appear to be testing what they are intended to test). VP based assessment seems most likely, of all forms of evaluation, to offer the best predictive validity, due to the integration of clinical decision making into the evaluation.  

Limited published evidence of validity exists. Initial work was not promising: Jerant and Azari (2004) attempted to validate the scores attained by medical students participating in web-based multimedia simulated patient cases (DxR Clinician™). The scores generated by the programme did not correlate with another measure of clinical reasoning ability, the Diagnostic Thinking Inventory, so concurrent validity was not established in this case.

Other work on validation of results from VP-based assessment is in progress, and includes recent work from the University of Ulm. A VP-based assessment was trialled on 147 medical students as a possible replacement for a written exam. The researchers concluded that while they demonstrated some evidence for its validity (primarily content validity), it was difficult to conclusively prove that it overall it was a valid means of assessment. This was mainly because establishing concurrent validity by direct comparison to other forms of assessment) was difficult. It has been proposed that VP-based tests measure different aspects of performance compared with other existing forms of assessment.

Despite the lack of evidence, VPs are still thought to be likely to become the most effective form of assessment of clinical competence. Already, a limited version of a VP-type assessment, a branched computerised case simulation programme, Primum® has been utilised as part of the high-stakes United States Medical Licensing Exams, as a section of the USMLE Step 3, (http://www.usmle.org/step-3/).

Regardless of their popularity, further rigorous evidence supporting their validity for assessment purposes is necessary before they are universally implemented.
1.6 Virtual reality programme utilised for the current project

1.6.1 Description of the Otago Virtual Hospital (OVH)

The OVH is a computer simulation programme that was developed primarily by University of Otago members of staff; Dr Phil Blyth and Judith Swan, (Faculty of Medicine) and Swee-Kin Loke (Higher Education Development Centre). It is an advanced form of a virtual patient, utilising virtual reality technology.

It is currently operating from Otago Medical School, Dunedin, New Zealand. The programme portrays a virtual emergency department and is designed to recreate a realistic representation of medical practice in this setting. To achieve this, the participants use avatars (‘a movable character that represents a person in a virtual reality environment’⁷¹) to act as doctors, patients or other members of staff. In the virtual environment students playing the role of junior emergency medicine doctors take a history from the ‘patient’ avatar via text chat, perform and interpret ‘examinations’, check vital signs and undertake bedside investigations. They are required to communicate differential diagnoses to a senior colleague, (another avatar who represents a medical registrar), before proceeding to order and review tests such as radiology or laboratory investigations. Participants are then expected to make decisions about appropriate treatment, including medications, fluids and possible admission to hospital. The participant is required to prescribe and document medical information as would be necessary in a real medical setting.

The OVH programme is based on the well-known original virtual world Second Life (http://secondlife.com/), which was built by Linden Lab in 2003. Rather than utilising Second Life itself, the OVH has instead been constructed on the New Zealand Virtual World Grid (NZVWG) which is an open access virtual world grid built with open source software, (http://nzvwg.org). The NZVWG has been developed mainly as an educational initiative and runs off New Zealand based servers.

---

The OVH concept was based on an existing model, the University of Auckland Virtual Medical Centre\textsuperscript{73}, however it differs from the Auckland version in that the patient is not pre-programmed. The patient is played in real-time by an online actor, (a standardized patient) so authentic human responses to questioning are provided, an approach supported by recent work by the Faculty of Medicine, Imperial College, London.\textsuperscript{57}

Results of tests and investigations are automated and where programmed, treatment can affect the patient’s physiological responses. However, the clinical case is open-ended, as the participant directs the entire pathway of the case, which allows, for example, revisiting of history, e.g. asking about allergies before prescribing.

1.6.2 Operating the OVH

Participants navigate the OVH programme using the keyboard arrow keys to move their avatar around the virtual emergency department. This avatar has been designed to resemble an archetypal emergency medicine junior doctor, and the gender of the avatar is altered to match each participant, in order to better contribute a sense of realism.

The clinical scenario is initiated with the participants being shown a triage form for the patient, before being instructed to proceed with standard patient management.

The patient avatar can be altered to resemble any gender or age group, and the patient’s body shape, skin colour and clothing can be characterised. The patient therefore is made to realistically represent the specific case. To communicate with the patient, the participants use typed chat, similar to instant messaging. The patient avatar is operated by an actor (who is concurrently online), so the ‘patient’ is able to respond to all questions in real-time, with content of the history dictated by a pre-scripted scenario \textit{(appendix 2)}. The patient can also move around the virtual emergency department if necessary, controlled by the actor.
Physical examinations are performed by mouse-clicking on the appropriate parts of the patient’s body. To obtain bedside tests such as blood pressure measurements, temperature and oxygen saturation, the participant needs to locate the appropriate equipment (all within the patient cubicle or nearby), and mouse-click to operate. A sphygmanometer, pulse oximeter, blood glucose meter, electrocardiography machine and urine dipstick strips are all located within or nearby the patient cubicle.

To order laboratory and radiology investigations, the participant is required to move their avatar back to the Emergency Department central station, locate and complete the necessary forms to request tests, before being able to look up results of investigations. Only standard test results such as blood counts and chest X-ray will become instantly available when ordered, and, for example, it is not possible to get an instant MRI or urine culture results. This is intended to better replicate a real medical environment.

The participant can initiate all standard medical treatments as required. This includes oxygen, intravenous fluids, blood and the prescription of a limited set of medications. Medications are limited to those readily available in a small emergency department, including analgesics, drugs for rate control, anticoagulant agents and common antibiotics.

Following treatment, a patient admission or discharge note is typed into a text document within the programme to replicate standard handover procedure in the hospital environment.

1.6.3 Current uses of the OVH

While still in the process of development, the OVH has been used for a case study of seven medical students, the findings of which suggested that in the virtual environment, students needed to learn to notice clinical salient features, as opposed to paper cases in which all given information is generally relevant.72
Figure 2. Otago Virtual Hospital: Triage form
Figure 3. Otago Virtual Hospital: Doctor taking text history from patient in emergency department cubicle
<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>Reference Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haemoglobin</td>
<td>113 g/L</td>
<td>115 - 155 g/L</td>
</tr>
<tr>
<td>HCT</td>
<td>0.39</td>
<td>0.35 - 0.46</td>
</tr>
<tr>
<td>MCV</td>
<td>92 fl</td>
<td>81 - 98 fl</td>
</tr>
<tr>
<td>MCH</td>
<td>31 pg</td>
<td>27.0 - 33.0 pg</td>
</tr>
<tr>
<td>Platelets</td>
<td>388 x 10^9/L</td>
<td>150 - 430 x 10^9/L</td>
</tr>
<tr>
<td>WBC</td>
<td>21.8</td>
<td>4.0 - 11.0 x 10^9/L</td>
</tr>
<tr>
<td>Neutrophils</td>
<td>19.2</td>
<td>1.9 - 7.5 x 10^9/L</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>0.6</td>
<td>0.9 - 4.0 x 10^9/L</td>
</tr>
<tr>
<td>Monocytes</td>
<td>0.2</td>
<td>0.2 - 1.0</td>
</tr>
<tr>
<td>Eosinophils</td>
<td>0.02</td>
<td>&lt; 0.51</td>
</tr>
<tr>
<td>Basophils</td>
<td>0.01</td>
<td>&lt; 0.21</td>
</tr>
<tr>
<td>TropT</td>
<td>0.003</td>
<td>&lt; 0.01 ug/L</td>
</tr>
</tbody>
</table>

Figure 4. Otago Virtual Hospital: Doctor obtaining virtual patient’s blood test results
1.7 Aims of the project

Using the OVH, the current study aims to provide initial evidence for the validity of the use of a virtual-reality based clinical case to assess clinical competence. Validity will be inferred from differences in performance between expert and novice cohorts. Performance will be evaluated by objective scoring of outcomes, and by qualitative assessment of differences in clinical reasoning ability.
Chapter Two: Methods

2.1 Participants

Participants were recruited from students currently enrolled at Otago Medical School in Dunedin, and from registered doctors residing in the Dunedin area. Ethical approval was provided by the University of Otago ethics committee (reference code 12/039, [appendix 1]). Written, signed informed consent was gained from all participants prior to undertaking the trial.

The participants comprised three study cohorts representing increasing levels of medical training and experience: early-training medical students (third-year students), late-training medical students (fifth-year students), and registered doctors. All participants were over 18 years of age. Prior to testing, participants provided limited personal data consisting of gender, age, ethnicity and stage of training.

Otago Medical School offers a six-year undergraduate degree course. Entry into the professional course occurs after the first year, and is followed by two years of primarily non-clinical training. The fourth and fifth years are clinical in nature, with final exams at the end of this period. The sixth year of training is known as a ‘trainee intern’ year, and consists mainly of clinical experience. Provisional registration with the New Zealand Medical Council is gained at graduation at the end of the six years. To attain general registration, a further year of accredited clinical work under supervision is required.

The early-training cohort consisted of four medical students who were in their third year of study, or who not yet started their fourth year of medical training. Two students in this cohort were Bachelor of Medical Science students (this entails a fourth year of study, but is research based and does not involve clinical training). None of these students had any significant clinical experience.
The late-training cohort was composed of 4 fifth-year medical students. These students had not yet completed their final exams, but had obtained at least one year of clinical experience in their training.

The registered doctor cohort comprised four qualified doctors, all of who held current New Zealand registration in a general or vocational scope. These participants included a house officer, general practice trainees, and a palliative medicine specialist. Their postgraduate experience ranged from one year of clinical experience to decades of clinical work.

2.2 Testing

Testing of the OVH simulated case was performed at the University of Otago, with all participants using University of Otago Medical school computers, (Macintosh operating system, OS X, version 10.8.2.) A selection of common internet browser programmes (Safari, Internet Explorer, Firefox) were available on the testing computers for use by the participants.

Participants were given a brief explanation of how to operate the virtual reality programme prior to testing, and the researcher was physically present during testing to provide limited technical assistance with the OVH programme if required. Assistance was limited to providing reminders on how to operate the navigation, examination, investigation and prescription commands. Participants were instructed to treat the virtual clinical case as if it were a real medical scenario, and to act as they would in real circumstances. As it was designed to recreate usual medical practice, participants were allowed to access resources that would be present in an emergency department such as online sites, and textbooks. The participants were given unlimited time to complete the scenario.

Two different actors were used for the testing. Both actors were experienced in playing medical scenarios, and were blinded to both the identity and stage of
training of the participants, and to the aims of the study.

The investigator was online throughout testing, present as an avatar representing a medical registrar. Following the initial history and examination stage, the investigator would request (via text chat) from the participant their differential diagnoses, with supporting clinical reasoning. All text chat and actions were recorded in text form copied directly from the OVH interface. This included the history taken from the patient, any conversations with the investigator, all examinations and investigations, as well as the final admission or discharge note.

The times taken to complete various aspects of the scenario were recorded. This included time taken to make diagnoses, and to complete the patient admission or discharge.

2.3 The clinical scenario

This scenario (appendix 2) was written by practising physicians, and was based on experiences with real patients. It was specifically designed to depict a vague and undifferentiated presentation, so effective diagnosis and management would require an adequate history, examination and investigation. The actors were directed to give out certain information in the history only if asked specifically, in order to better represent a genuine medically naive patient who would be unaware of salient points important for diagnosis.

The virtual patient (known as Gertrude MacFarlane) is represented by an avatar of an elderly woman, initially located seated in a cubicle within the emergency department. She is a 76-year-old woman, presenting with confusion and complaining of feeling generally unwell. Underlying the vague presentation is a five-day history of symptoms consistent with a urinary tract infection, poor medication adherence, with consequent fast atrial fibrillation, as well as dehydration. The case is intended to represent a scenario with a level of difficulty that would be easily
diagnosed and managed by a graduate doctor within their first year of independent clinical work.

2.4 Analysis

2.4.1 Quantitative scoring of overall performance

A quantitative scorecard to measure overall performance in the clinical case was developed from a consensus of expert opinion prior to testing. This was in keeping with the latest recommendations from the most recent consensus statement for technology-enabled assessment of health professions education, (Ottawa Conference 2010): ‘Researchers should develop scoring methods that automate the collection, integration and analysis of the vast and often novel information available through technology enabled assessment.’

The Clinical Performance Scorecard appears in figure 5.

The scorecard was based on a similar tool developed by Botezatu et al, designed specifically to assess performance in a virtual patient clinical case. Scoring incorporated the participant’s ability to elicit important aspects from the history, examination and investigations, to make correct the diagnoses and list adequate differentials, as well as including a measure of achievement of expected management decisions. In keeping with the clinical reasoning literature, there was no correct pathway through the problem, achievement of the outcomes could occur in almost any order.

The design of the scorecard also took into account some aspects of underlying theory of key features type problems, in that outcomes assessed were only those that were necessary to resolve the case. As the case was designed to be a poorly differentiated clinical scenario, and the participant was required to elicit the history without guidance, this resulted in considerably more elements involved than a usual key features problem.
Over-investigation, dangerous omissions or treatment decisions resulted in penalty scores, built into the scoring system.

Comparison of scores between cohorts was performed using one-way ANOVA with a Bonferroni test used post-hoc, using statistical package Stata, (Stata 12.1/IC for Mac, StataCorp LP, College Station, TX).
<table>
<thead>
<tr>
<th>Item</th>
<th>Positive Score</th>
<th>Negative Score</th>
<th>Maximum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient History</strong></td>
<td>All relevant aspects obtained:</td>
<td>+1</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>PC: brought in by neighbour with confusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Last few days history of:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Tired/feeling unwell</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Short of breath on exertion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Not eating or drinking as usual</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Not taken medications</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Not passing much urine/smelly urine</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Some relevant negatives: no chest pain, no cough</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Med Hx:</td>
<td>apparently fit and well</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meds:</td>
<td>betaloc, cartia, no allergies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soc Hx:</td>
<td>lives alone, independently, nil alcohol or smoking history</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical Examination</strong></td>
<td>Assessed all of:</td>
<td>+1</td>
<td>Missed one of these aspects</td>
</tr>
<tr>
<td></td>
<td>• Blood pressure, pulse, oxygen saturation, temperature, blood glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Chest exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Abdomen exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Investigations</strong></td>
<td>All of listed and no more than 10% extra:</td>
<td>+2</td>
<td>More than 20% unnecessary</td>
</tr>
<tr>
<td></td>
<td>Diagnosis diagnosis of AF and UTI</td>
<td></td>
<td>Missed critical investigations:</td>
</tr>
<tr>
<td></td>
<td>Diagnosis of co-existing conditions:</td>
<td></td>
<td>• ECG, urine</td>
</tr>
<tr>
<td></td>
<td>Adequate differential diagnoses:</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• includes chest infection, MI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td>Correct diagnosis of AF and UTI</td>
<td>+1</td>
<td>Potentially life-threatening decision</td>
</tr>
<tr>
<td></td>
<td>Diagnosis of co-existing conditions:</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• dehydration &amp; confusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adequate differential diagnoses:</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• includes chest infection, MI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Management</strong></td>
<td>Initial management:</td>
<td>+0.5</td>
<td>Potentially life-threatening decision</td>
</tr>
<tr>
<td></td>
<td>• Low flow oxygen</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• venous access</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• suitable rehydration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Finding out doses of usual meds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate treatment for positive diagnosis:</td>
<td></td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Rate control – metoprolol or similar</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Anticoagulation – warfarin discussed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Suitable antimicrobial for UTI – e.g trimethoprim</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate follow-up:</td>
<td></td>
<td>+0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Monitoring of rate and hydration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Admission or clear discharge plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Clinical performance scorecard
2.4.2 Sub-score assessment and analysis

Measured separately was the achievement of complete and correct diagnoses; to be used as an overall indication of diagnostic ability. Achievement of safe effective management was also measured as a separate result, as this was presumed to best reflect the outcomes for the patient.

Comparison of scores between cohorts were analysed using Fishers Exact test.

2.4.3 Quantitative analysis of times taken

Times taken to make the diagnoses were measured, as were times taken to complete the clinical case. These were intended as a measure of efficiency. Statistical analysis was again by one-way ANOVA with a Bonferroni test used post-hoc, using statistical package Stata 12.1 (Stata 12.1/IC for Mac, StataCorp LP, College Station, TX).

2.4.4 Qualitative analysis

A general inductive approach was utilised to identify key themes from within the data. The inductive analysis involved deriving concepts or themes that repeatedly emerged from the raw transcript data, based on the evaluation objectives. Themes relating to clinical reasoning were identified and these included analysis of the pathway to the diagnoses, (as an indicator of the type of reasoning process), analysis of diagnostic accuracy and also the nature of any errors.

A general inductive approach was also used for analysis of the summary documents produced by the participants at the conclusion of the case. Analysis was undertaken to identify any apparent differences between cohorts in the structure or content of the notes.
Chapter Three: Results

3.1 Quantitative analysis

3.1.1 Demographic data of participants

Twelve participants were recruited from the University of Otago; these included both employees and students of the institution. Demographic details of the participants are displayed in table 1. Participants were of mixed ethnicities, and ages increased with stage of training. The majority of participants were male.

Table 1. Demographic details of participants

<table>
<thead>
<tr>
<th>Stage of training</th>
<th>Average age (± 1 S.D.)</th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical student: third-year (n=4)</td>
<td>21.25 ± 1.26 years</td>
<td>2 male 2 female</td>
<td>1 NZ European  2 Middle Eastern/Arabic  1 Asian</td>
</tr>
<tr>
<td>Medical student: fifth-year (n=4)</td>
<td>25.5 ± 4.43 years</td>
<td>4 male</td>
<td>1 NZ European  1 Middle Eastern  1 South African  1 NZ Chinese</td>
</tr>
<tr>
<td>Registered doctors (n=4)</td>
<td>34.5 ± 11.03 years</td>
<td>3 male 1 female</td>
<td>2 NZ European  1 Middle Eastern  1 NZ/East Slav</td>
</tr>
</tbody>
</table>

3.1.2 Overall scores achieved

The overall score for performance in the virtual clinical case was measured using the Clinical Performance Scorecard, which was developed from collective expert opinion prior to testing, (figure 5).
The doctor cohort achieved the highest mean scores (9.125 out of a possible 10), and this was significantly higher \((p< 0.05)\) than the mean score of the third-year student cohort (4.75). The doctor cohort also achieved a higher score than the fifth-year cohort (6.5), but this did not reach the level of significance. The fifth-year students achieved a higher score than the early training third-year cohort, but this was also not significant. Results are displayed in table 2.

**Table 2.1. Mean scores of clinical performance (out of possible 10)**

<table>
<thead>
<tr>
<th>Stage of Training</th>
<th>Mean score (± 1 S.D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-year students (n=4)</td>
<td>4.75 ± 2.72</td>
</tr>
<tr>
<td>Fifth-year students (n=4)</td>
<td>6.5 ± 1.34</td>
</tr>
<tr>
<td>Qualified doctors (n=4)</td>
<td>9.125 ± 0.66</td>
</tr>
</tbody>
</table>

**Table 2.2. Significance of differences between clinical performance scores**

<table>
<thead>
<tr>
<th></th>
<th>Third-year students</th>
<th>Fifth-year students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fifth-year</td>
<td>p= 0.042</td>
<td>N/A</td>
</tr>
<tr>
<td>Doctor</td>
<td>p = 0.022*</td>
<td>p=0.281</td>
</tr>
</tbody>
</table>

Some individual aspects within the scorecard were analysed separately. The achievement of the correct and complete diagnoses was assessed alone, as a measure of diagnostic ability, regardless of clinical behaviour. Achievement of correct management was also assessed, as a measure of the key outcomes for the patient.

**3.1.3 Achievement of complete diagnoses**

Within the virtual case, participants were required to diagnose the patient with all three key clinical issues: probable urinary tract infection (with associated confusion), fast atrial fibrillation, and dehydration. Their ability to do this was assessed by both
asking the participants for diagnoses within the case itself, and their written record of diagnoses within the admission or discharge note.

All of the participants from the doctor cohort were able to make the correct and complete diagnoses, however only two participants from the fifth-year cohort achieved this. None of the third-year students were able to make complete diagnoses for this case (table 3).

**Table 3. Achievement of complete diagnoses by stage of training**

<table>
<thead>
<tr>
<th>Stage of training</th>
<th>Complete diagnoses achieved</th>
<th>Not achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-year students</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Fifth-year students</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Doctors</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Significance was measured using Fishers Exact Test, and indicates that stage of training and ability to make a complete diagnosis in this case are significantly related (p<0.05).

### 3.1.4 Achievement of safe effective management

For participants to achieve safe effective management within the virtual case they were required to achieve four key objectives: to safely prescribe a rate control agent for the atrial fibrillation, to prescribe a suitable anti-microbial medication for the infection, arrange some form of rehydration, and make adequate documented arrangements for admission or discharge.

Prescriptions of medication and provision of fluids was assessed from the case transcripts, and admission plans were evident from within the participants’ written admission or discharge notes.
Results are displayed in table 4. Again, all of the registered doctors achieved the management objectives, but only a single fifth-year student managed to do so and none of the third-year student cohort was able to achieve the desired outcome objectives.

Table 4. Achievement of safe effective management by stage of training

<table>
<thead>
<tr>
<th>Stage of training</th>
<th>Safe effective management achieved</th>
<th>Not achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-year students</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Fifth-year students</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Doctors</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Relationship between stage of training and ability to achieve safe effective management within the virtual case was again assessed using Fishers Exact Test, and was significant (p<0.05).

3.1.5 Time taken to make diagnoses

The time taken to make the diagnoses was recorded for all participants, regardless of whether the diagnosis was correct or complete. Results are shown in table 5.

The doctor cohort was the fastest to make their diagnoses with a mean time of 49 minutes, compared to the fifth-year student cohorts (69 minutes) and the third-year students (also 69 minutes). However, this did not reach the level of statistical significance (analysed by one-way ANOVA) due to the low numbers of participants. There was no difference in mean times between the third and fifth-year student cohorts.
3.1.6 Time taken to achieve correct and full diagnoses

Measurement of the time taken to make correct and complete diagnoses was also measured. Results are shown in table 5.

All participants in the doctor cohort achieved the full correct diagnoses and this was with a mean time of 49 minutes. Only two participants from the fifth-year group achieved the complete diagnoses, and this was achieved with a mean time of 70 minutes. No participants from the third-year medical student cohort achieved complete diagnoses in the clinical case. One-way ANOVA did not show any significant difference between these times. There were insufficient measurements recorded to test for trend.

Table 5. Times taken to make diagnoses

<table>
<thead>
<tr>
<th>Stage of training</th>
<th>Mean time to make any diagnosis (± 1 SD)</th>
<th>Mean time to full correct diagnosis (± 1 SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-year students</td>
<td>69 ± 22.4 minutes (n=4)</td>
<td>not achieved</td>
</tr>
<tr>
<td>Fifth-year students</td>
<td>69 ± 9.6 minutes (n=4)</td>
<td>70 ± 15.6 minutes (n=2)</td>
</tr>
<tr>
<td>Doctors</td>
<td>49 ± 7.1 minutes (n=4)</td>
<td>49 ± 7.1 minutes (n=4)</td>
</tr>
</tbody>
</table>

3.1.7 Time taken to complete clinical case

The time taken to complete the clinical case was measured for all participants. Results are shown in table 6.

The doctor cohort completed the case (regardless of outcomes) in the shortest time (90.8 minutes), compared with both student cohorts. Fifth-year students took the longest time to complete the case, with a mean time of 111.5 minutes, compared with third-year students who finished in a mean time of 107 minutes. There were no significant differences between these times.
To complete the case safely, the participants were required to satisfy the clinical objectives for safe effective management, (as stated above: rate control, antimicrobial treatment, rehydration, admission plan.) All doctors were able to safely meet the clinical objectives within the mean times recorded. None of the third-year students was able to achieve these objectives, and only one fifth-year student did so; this participant completed the case in 79 minutes, which was quicker than the mean time recorded by the doctor cohort. Results are in table 6.

Again, however, ANOVA indicated no significant difference in times taken between the groups.

**Table 6. Times taken to complete clinical case**

<table>
<thead>
<tr>
<th>Stage of training</th>
<th>Time to finish case (± 1 S.D.)</th>
<th>Time to complete case with full safe management (± 1 S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-year students</td>
<td>107 ± 27.0 minutes (n=4)</td>
<td>Not achieved</td>
</tr>
<tr>
<td>Fifth-year students</td>
<td>111.5 ± 27.4 minutes (n=4)</td>
<td>79 min (n=1)</td>
</tr>
<tr>
<td>Doctors</td>
<td>90.8 ± 9.0 minutes (n=4)</td>
<td>90.8 ± 9.0 minutes (n=4)</td>
</tr>
</tbody>
</table>
3.2 Qualitative analysis

3.2.1 Thematic analysis of case transcripts

All participants’ written conversations and actions within the virtual clinical case were recorded electronically as a text document. To perform qualitative analysis of the transcripts, a general inductive approach was utilised to identify key themes from within the data. The inductive analysis involved deriving concepts or themes that repeatedly emerged from the raw data, based on the evaluation objectives.

One of the aims of the study was to evaluate the effectiveness of the virtual case as a tool to assess clinical reasoning skills. For the purposes of this analysis, themes relating to clinical reasoning were identified. Based on the literature, three main indicators of clinical reasoning ability were chosen. These themes were comprised of the pathway taken to the diagnoses (as an indicator of the efficiency of the reasoning process), the accuracy of the diagnoses, representing the outcomes of clinical reasoning, and the nature of any errors.

During the analysis it became apparent that assessment of diagnostic processes and outcomes explained only part of the differences between the cohorts in their clinical reasoning capabilities. The ability to transform presenting information into key clinical concepts also appeared to be important. This was the ability to identify key features of the clinical case, a skill that incorporated not only diagnostic ability, but also knowledge and application of management decisions.

Because clinical reasoning is described in different ways, some of the information could be interpreted to correspond with more than one theme.

In the following chapter, all transcripts have been copied verbatim, with no correction of typing and spelling errors. However, to maintain clarity, the participants have been coded anonymously by their stage of training, and numbered in the order of participation.
3.2.1.1 Theme 1: Efficient ‘pattern recognition’ and the formation of illness scripts

The pathway taken by participants to make the main diagnosis illustrated this theme.

Qualitative analysis of the case transcripts indicated that with more clinical experience, the main diagnosis of urinary tract infection (UTI) was made with more efficiency, via a more direct pathway. It appeared that subjects with more experience could quickly generate the correct hypothesis earlier in the clinical case. This appeared to be because they more rapidly recognised the ‘pattern’ of the initial presentation. They were then able to confirm their diagnosis by efficiently matching the pattern to underlying illness scripts, (schemas of existing knowledge about a disease process and its presentation.)

This is consistent with the development of non-analytical clinical reasoning skills. Underlying non-analytical clinical reasoning is the ability to form and retrieve relevant illness scripts. It was evident that with increasing competence, the correct diagnosis was obtained more efficiently due to the ability to match the ‘fit’ of the presenting information to these previously formed illness scripts. The existence of a well-organised and defined illness script resulted in a more efficient pathway to diagnosis.

(Obviously effective illness script activation and assessment also contributed to the accuracy of diagnoses, however this was analysed separately with respect to development of elaborated knowledge.)

Evidence from the transcripts:

The presentation in this virtual case was designed to illustrate a common clinical situation in which underlying infection may present as confusion in an elderly patient. The virtual patient in the clinical case presented only with confusion, and
feeling generally unwell. She did not present specifically with urinary tract symptoms, unless asked directly.

Despite the non-specific presentation, all participants from the registered doctor cohort appeared able to generate the diagnosis of UTI from the initial presenting information in the history. It appeared that these experienced participants recognised the pattern of *‘an elderly woman with confusion, mild fever, and no other indicators of infection’* as being most consistent with a UTI. They all then confirmed their diagnosis with a urinary dipstick. While most also sought further information in the history about urinary symptoms, one participant deliberately omitted to do so, justifying that absence of symptoms would not much alter the likelihood of UTI still being a probable cause of her confusion.

The direct path to diagnosis is illustrated by this sample of transcript from one medically qualified participant, recorded 11 minutes into the scenario. This participant was in the process of taking the history of the presenting problem.

**[Doctor 1]** Can you tell me, has anything else felt wrong this week? Apart from feeling tired and short of breath?

**[Gertrude Macfarlane]** let me think
i wasn’t eating or drinking as usual
but now i am very very thirsty
i usually don’t feel like this

**[Doctor 1]** And it’s just been the last few days?

**[Gertrude Macfarlane]** yes

**[Doctor 1]** Has it been the same, or getting worse?

**Gertrude Macfarlane:** i don’t know
[Doctor 1] ok

[Gertrude Macfarlane] i think getting worse
i think i have slight fever too
and my neighbour says i was confused
well, don't know why she felt so
ooh i remember, she said I haven’t fed my cat, my poor kitty :(

[irrelevant chat about patient’s cat]

[Doctor 1] Well hopefully, but we’d better figure out what's going on first

[Gertrude Macfarlane] ok

[Doctor 1] Have you had any pain? Are you sore anywhere?

[Gertrude Macfarlane] no

[Doctor 1] Have you been coughing?

[Gertrude Macfarlane] nopes

[Doctor 1] Any problems with your waterworks? Are you passing more or less urine, or does it sting at all?

[Gertrude Macfarlane] ooh yes, i forgot to tell you about it
i haven't been going to toilet as usual
and last time it really stank!!

This participant asked directly about urinary symptoms while still exploring the initial presenting problems of fatigue, confusion and dyspnoea. The participant appears to
be generating possible diagnoses for the presenting issues, and it can be inferred from the questions that they have considered options such as a respiratory tract infection (‘have you been coughing’) and UTI (‘any problems with your waterworks?’).

Other participants from the medically qualified cohort showed similar clinical behaviour. This demonstrates the concept that clinically experienced doctors had well organised underlying illness scripts relating to this presentation. They all rapidly recognised that the presentation of an elderly woman with acute onset of confusion could be due to a possible underlying UTI. The efficient retrieval of this relevant illness script was most likely due to their more extensive clinical experience. These participants would have had multiple previous examples of similar presentations (exemplars) with which to compare the current case.

The participants from the fifth-year student cohort were also able to make the diagnosis of UTI, but could do so usually after taking a full history, examination and performing a urine dipstick, rather than recognising the diagnosis from the initial presentation. The fifth-year students usually took longer to make the diagnosis, and seemed to require more information to do so. Their histories were more structured and they sometimes gained more irrelevant information. This group was more likely to ask about urinary symptoms later in the history, or as part of a review of systems, rather than as directly relevant to the presenting complaint.

This suggests that their illness scripts were less accessible and not so clearly organised.

To illustrate is a portion of a transcript from a fifth-year student: recorded 45 minutes into the scenario. The majority of the medical history, encompassing the presenting problem, medications, family and social history, preceded it.

[Fifth-year student 4] is there anything else you thought you should mention, that i havent already covered?
[Gertrude Macfarlane] Ummm... No not really

[Fifth-year student 4] okay. just a few more questions. have you noticed any changes in your bowel habit?

[Gertrude Macfarlane] No my bowls have been fine

[Fifth-year student 4] okay. and the waterworks?

[Gertrude Macfarlane] Oh well actually yes, they've been a bit odd
I haven't been peeing as much as usual
And I think it smells a bit funny

[Fifth-year student 4] right. has the colour changed at all?

This indicates that the diagnosis of UTI had not occurred to this participant while exploring the presenting issues. The urinary symptoms only emerged in response to questioning when the participant was systematically reviewing body systems at the end of the medical history. This suggests that this particular participant was not yet comparing the patient’s presentation to an underlying illness script, but instead was still trying to generate the initial diagnosis, using the review of systems approach. The pathway to the diagnosis was obviously much less direct than that of the doctors’.

Only some of the third-year students managed to make the key diagnosis of UTI. If they did so, this appeared to be as a consequence of recognising the significance of the findings of an incidental urinary dipstick, or as a result of a structured review of symptoms in which they asked about changes in urinary patterns. It did not appear that UTI was a differential diagnosis for the initial presentation, but instead that the third-year students seemed to ‘stumble upon’ the correct diagnosis. However most participants in this group did manage to identify the presence of some kind of
infection, noting primarily the new onset of the presenting symptoms of fever and confusion, (although some did not recognise 37.8°C as febrile.) The diagnosis of UTI usually did not occur until towards the end of the clinical case, if at all.

This is demonstrated by a segment of transcript recorded 56 minutes into the scenario, following the history and examination. At this stage a urinary dipstick had not been requested.

[Registrar] can you please tell me what you think is wrong with Mrs McF and give me your main differential diagnoses, and reasons :) 

[Third-year student 2] Infection because confusion in elderly, fever, loss of appetite and fatigue and tachycardia due to discontinuation of beta blockers. Only other possibilities are endocrine issues or a malignancy.

[Registrar] ok, what is the most likely in your opinion  

[Third-year student 2] infection of some sort  
because it's quite sudden

It appeared as though this participant had not yet specifically considered UTI as a diagnosis, although they had at this stage gained similar information from the history and examination as all the other participants. Similar clinical behaviour was observed in the other third-year student participants, which suggested that they did not follow any clear pathway to the diagnosis, so possibly have not yet developed any comprehensive illness scripts at all. Due to their underlying pathophysiological knowledge, they sometimes could make the diagnosis of UTI from investigation findings alone.

Overall, the cohorts generally exhibited different clinical behaviours that were evident from the pathways chosen to make the diagnosis of UTI. With increasing clinical experience (from which greater competence can be presumed), there was
evidence of a more direct and efficient route taken to the diagnosis. This is consistent with current clinical reasoning theory; that expertise involves development of a case-specific illness scripts.

3.2.1.2 Theme 2: Elaborated knowledge

Elaborated knowledge is the ability to retrieve and fit existing knowledge to a new clinical scenario. It involves the ability to organise the presenting information and then examine all presenting features in combination, comparing it with underlying schemas. Elaborated knowledge results in better diagnostic ability.\textsuperscript{14}

3.2.1.2a Diagnostic accuracy as evidence of elaborated knowledge:

Skilful clinical reasoning, by definition, results in diagnostic expertise - the generation of accurate, likely and unified diagnoses. In this analysis, overall diagnostic accuracy was taken as evidence of effective elaboration of knowledge, providing an indication of the ability to retrieve and apply knowledge to a new problem. As part of this, the presence of diagnostic errors was also evaluated.

It should be noted that diagnostic accuracy was not unequivocally an indication of clinical reasoning expertise, as it was not an evaluation of the underlying process involved in obtaining the diagnosis. It was possible, although less likely, to make the correct diagnosis using a poor or invalid reasoning process.

Analysis of the transcripts indicated that participants from the qualified doctor cohort consistently produced the most accurate and likely diagnoses.

Accuracy of diagnosis was assessed by the ‘best fit’ to the presenting symptoms. Participants from the medically qualified group were usually able to identify relevant presenting issues to generate their diagnoses, and symptoms were evaluated in
combination rather than isolation. The doctors sought diagnoses that encompassed all of the presenting features, and seemed able to better compare and fit these features to their underlying knowledge, i.e. they elaborated their knowledge. The diagnoses suggested were all more commonly encountered clinical problems - more probable diagnoses, again reflecting greater clinical experience.

All of the doctors made the primary diagnosis of UTI, and when requested, could justify this diagnosis. They generally did so by linking the clinical presentation with the relevant investigation findings, although usually quite succinctly:

This is evident from these transcripts, showing the justifications for the diagnosis of UTI:

[Doctor 3] probable UTI... dipstick urine is positive for leucocytes and protein+++  

[Doctor 4] underlying infection..urinary.. febrile ..has had some symptoms  

[Doctor 2] UTI: she noticed a smell with her urine and urine dipstick also confirmed nitrates and protein in her urine. She has a low grade fever for which I don't have any other source!  

Differential diagnoses for the presenting symptoms of fatigue and confusion included other common sources of infections, including:

[Doctor 4] maybe a LRTI  

Differential diagnoses for the tachycardia were also suggested. While the participants in this group noted that the concurrent infection and non-adherence to rate-control medication were the likely cause of the tachycardia/exacerbation of AF, (discussed further in section 3), they were also able to generate feasible alternative reasons for the cardiac issues. The diagnoses proposed included cardiac ischaemia or other concurrent illness, both reasonably likely in the given clinical scenario.
Doctor 2] Ischaemic causes

Doctor 4] has some rate related ischaemia on her ecg

Doctor 3] any incidental illness

Medically qualified participants were also able to suggest other possible diagnoses for the new onset of confusion, again offering only diagnoses that were likely in this clinical setting. These included:

Doctor 3] TIA

Doctor 2] I thought she has the risk factors for an embolic CVA... could that cause the confusion?

The medically qualified participants mainly suggested diagnoses that took into account all of the presenting problems, and they were diagnoses that were likely in the context of the clinical presentation. The participants’ evaluation of all the presenting issues demonstrated their elaborated knowledge. The succinct explanations for the diagnosis of the UTI suggests that their knowledge at times may have become compiled, i.e. knowledge that was present but did not need to be explained.

By comparison, fifth-year students were more likely to generate diagnoses that were wider but less accurate. They more often proposed diagnoses based on individual symptoms or signs in isolation, demonstrating knowledge that was more dispersed than elaborated. They were more likely to offer rare diagnoses.

Reflecting this, although all fifth-year participants were eventually able to make the main diagnosis of UTI, one participant also suggested a less likely primary diagnosis of nephrotic syndrome, based on the observations of proteinuria and dyspnoea.
Similar to the participants from the doctor cohort, all participants from the fifth-year cohort were (eventually) able to combine relevant factors from the history with the examination findings to justify the diagnosis of UTI.

Justifications given for the diagnosis of UTI from transcripts included:

[Fifth-year student 4] a background of one week of confusion and generalised fatigue and inability to take care of herself and eat or drink. She has a temperature of 38.2, and has reduced UO with proteinurea and raised WCC. So i wonder if she has some underlying infection

[Registrar] what kind of infection were you thinking

[Fifth-year student 4] well i thought urinary infection considering her di[stick result.

Other fifth-year students suggested similar explanations:

[Fifth-year student 1] well I think she has got a UTI ... this is because her urine dipstick is positive for WBC and she has got a Temp of 38.1

[Fifth-year student 2] infection as she is tachycardic with a temperature of 38.2 degrees celcius and she also mentioned feeling feverish yesterday. She also has pain on urination and strong smelling urine which is positive on the dipstick for leucocytes, protein and nitrites.

These diagnoses were quite well elaborated.

The differential diagnoses proposed by this group appeared much wider and less likely than those from the doctor cohort. Proteinuria seemed to be a strong focus for differential generation, with nephrotic syndrome, pyelonephritis and nephritis all
suggested. This may well have been due to participants noting that the project’s supervising physician was a renal consultant.

This is shown in the following examples:

[Fifth-year student 3]..I think she may be developing nephrotic syndrome..
The protein in her urine makes me suspicious
Also the history of shortness of breath on exertion and waking up at night short of breath makes me think that she may be retaining fluid

[Fifth-year student 2] possibly upper uti with pyelonephritis.

[Fifth-year student 1]…bactermia [sic] of some sort, and rare things like nephritis

The differential diagnoses proposed were generally a poorer fit to the clinical information. An example of this was the differential diagnosis of tuberculosis, which was unlikely in an elderly New Zealand woman with no history of travel, exposure, or respiratory symptoms.

[Fifth-year student 3] Infection …Given her night sweats and temperature, I would guess bacterial. Maybe TB..

Key clinical findings were absent for many of the proposed differential diagnoses – for example, nephritis was not a likely diagnosis in the absence of haematuria, and the patient with nephrotic syndrome would have presented with oedema evident.

Some of the differential diagnoses suggested included more rarely encountered clinical conditions. This is evident from the much wider range of infections suggested:

[Fifth-year student 2] pyelonephritis.

[Fifth-year student 4] probably had some sort of infection (resp...

Similar to the doctor cohort, participants from this group were also able to suggest some other feasible causes for the confusion:

[Fifth-year student 7] A TIA could present similarly

Occasionally, some other causes were also suggested for the tachycardia:

[Fifth-year student 7] but it could be a cardiac insult

In general, the fifth-year students showed less diagnostic accuracy than the doctors, reflecting less elaborated knowledge.

The third-year students overall exhibited poor diagnostic ability, obviously reflecting their lack of clinical experience. Their knowledge appeared restricted, relating purely to biomedical knowledge and could not be elaborated to the patient presentation.

The differential diagnoses produced by the third-year student cohort were very broad and often unlikely, and generally were a poor match to the patient presentation. The diagnoses proposed appeared to more often reflect local third-year block module teaching (i.e. endocrine, cancer, neurology and renal) rather than a match to the patient’s presenting symptoms.

While most participants from the third-year group did manage to produce the primary diagnosis of UTI, the evidence supplied for the diagnosis was missing or poorly described. Some examples from the transcripts follows:
[**Third-year student 2**] Infection because confusion in elderly, fever, loss of apetite and fatigue

[**Third-year student 1**] all her symptoms points to Hypovolaemia, and an infection might be the underlying causes ..she have proteinuria so thats brings us to urinary tract infections’

There were no obvious attempts to generate diagnoses that could unify the presenting complaints. The third-year students mostly diagnosed each presenting problem separately, with a pathology relating only to that system. Illustrating this is the diagnosis of post-streptococcal glomerulonephritis to explain the proteinuria:

[**Third-year student 4**] that pos-strep GN might explain the positive protein in her urine

Suggestions for causes of dehydration included diabetes, (assumed to mean diabetes mellitus):

[**Third-year student 3**] diabetes due to her increased thirst and urination presentations..

Confusion was sometimes considered to be a purely neurological disorder, illustrated by these transcripts:

[**Third-year student 3**] dementia and alzheimers - given her age and presentation of confusion and forgetfulness, although considering the length of there symptoms that seems unlikely

Diagnoses to explain the fatigue and tachycardia were extremely wide, non-specific and less probable. They included thyroid dysfunction and cancer.
[Third-year student 3] potentially thyroid issues due to her tiredness and tachycardia.

[Third-year student 2] Fatigue and loss of aptitude is the main symptoms that may lead to the endocrine/malignancy conclusion.

It is evident that the third-year students often suggested diagnoses that had key symptoms absent. The following statement demonstrates this:

[Third-year student 4] Yes, I was thinking of pneumonia as well. Could explain why she feels sleepy all the time.

Pneumonia was not a likely diagnosis given that the patient had no respiratory symptoms, an unremarkable chest examination, and a normal chest x-ray.

It appeared that the third-year students were obviously unable to elaborate their diagnoses effectively, exhibiting knowledge that was dispersed at best.

One participant made an isolated diagnosis of angina. This was the result of obtaining a history from the virtual patient of previous similar episodes of dyspnoea, without exploring any further symptoms. This illustrated the diagnostic error of premature closure, i.e. making a diagnosis without seeking further information to support or refute the initial hypothesis. They justified the diagnosis as follows:

[Third-year student 1] well pulse is v high ...and the ECG also indicate fast beating of the heart... From the history... she had similar episodes of the same symptoms.

In summary, with increasing competence there was evidence of better diagnostic accuracy. With greater clinical experience, participants produced diagnoses that were more accurate and more probable. The diagnostic accuracy reflected the development of elaborated knowledge - the ability to apply existing knowledge to a new clinical problem. Consistent with the development of elaborated knowledge, it
was evident that more experienced participants were better able to recognise the important presenting issues, and evaluate them in combination to produce their diagnoses. Participants from both cohorts of medical students instead often produced diagnoses from isolated symptoms, demonstrating dispersed knowledge. The third-year students, with no clinical experience, were often unable to generate any feasible diagnoses for some of the presenting problems. This indicated that while still in the early stages of medical training, their underlying knowledge structure was largely biomedical in nature.

3.2.1.2b The ability to support and integrate diagnoses:

Also essential to elaborated knowledge is the ability to integrate and explain all presenting clinical issues. A theme that emerged from the transcripts was the difference in clinical behaviour pertaining to the recognition of the patient’s dehydration.

To diagnose the patient’s dehydration in the virtual case, the participants were required to recognise relevant aspects from the history and from the examination findings. The virtual patient in the clinical case volunteered in the history that she was thirsty, and that her oral intake had been poor. On examination, she was clinically dehydrated with dry mucous membranes and a jugular venous pressure of 0 cm. Other potentially relevant findings included tachycardia (although in fast AF), and normal blood pressure.

It was expected that a competent participant would note the significance of the clinical findings such as dry mucous membranes, and could identify that in this setting, the dehydration was most likely due to the patient not consuming enough fluid as a result of her infection-induced confusion.

While most participants were able to make the diagnosis of dehydration, there was a difference between the cohorts in the evidence used to support the diagnosis, with
the qualified doctors more likely to justify the diagnosis of dehydration with specific clinical evidence such as dry mucous membranes. Notably, it was primarily members of the doctor cohort who connected the patient’s dehydration with her other clinical issues, explaining specifically why the dehydration had occurred.

This indicated that the medically qualified doctors were the only participants able to form links between all of the diagnoses. Participants from this group were able to specifically connect the dehydration to the confusion that was a consequence of the concurrent infection.

To illustrate is an example from a transcript from one of the medically qualified participants:

[Doctor 1] She’s also dehydrated…. just from not drinking enough from confusion/unwell

[Registrar] I see

[Doctor 1]: I should give her some fluids

Later in this participant’s admission note, they also specifically listed amongst the relevant findings:

‘JVP low, tongue dry.’

So from this it appeared that this participant was able to propose a reason for the dehydration, and to justify the diagnosis with the relevant specific clinical findings.

Below is another example from a participant from the doctor cohort:

[Doctor 3] I’ve got a delightful 76 yr old lady
She has known prev cardiac disease and is on aspirin and metoprolol for what sound like AF

[Registrar] sure

[Doctor 3] She presents with a few days of confusion (mild) and significant reduction of ADLs

[Registrar] why do you think that is?

[Doctor 3] The significant findings are: irreg irreg tachycardia of 132, tachypnoea of 26, dehydration (JVP normal, dry tongue), chest clear, abdo NAD dipstick urine is positive for leucocytes and protein++, blood glucose normal
My working diagnosis is: confusion and exacerbation of AF associated with probable UTI. Other possibilities are not taking medication or any other incidental illness. A TIA for instance.

This illustrates that this participant was able to unify and elaborate their diagnoses. They supported the diagnosis of dehydration with clinical findings, and then linked together the diagnoses of UTI and AF.

Participants from the fifth-year student cohort exhibited slightly different clinical behaviour with respect to the patient’s dehydration. Most noted the dehydration, and this was evident from the following comments:

[Fifth-year student 1] She is volume depleted and tachy.

[Fifth-year student 2] She does seem quite dry

[Fifth-year student 4] this is a 76 year old lady who is dehydrated
However, the fifth-year students only occasionally elaborated on any specific clinical evidence for dehydration, and there were rarely explanations offered on why the dehydration had occurred.

Only some of the third-year students suggested a diagnosis of dehydration. When clinical evidence was given for this diagnosis, participants attributed clinical evidence such as the elevated heart rate to the dehydration without consideration of other possible reasons for these less specific signs. This is demonstrated below.

**[Third-year student 4]** So she have symptoms of dehydrations, complains of feeling thirsty today, her heart rate was raised and her BP was low.

It would appear that this participant made the diagnosis of dehydration based on the tachycardia, and from an interpretation of a normotensive blood pressure as hypotensive. No participants in this group noted specific signs of dehydration such as dry mucous membranes.

Only one participant from the third-year student cohort attempted to link the dehydration to other presenting problems by offering this explanation:

**[Third-year student 4]** I think Mrs Mc Farlane might have hypovolaemia secondary to an infection

However, this connection was not able to be further explained.

In summary, only the qualified doctors were consistently shown to support their diagnosis of dehydration with specific clinical findings, and connect this diagnosis to the other presenting issues. This shows further evidence of the development of elaborated knowledge, which involves the ability to unify and elaborate diagnoses.
3.2.1.3 Theme 3: Recognition of key features

While ‘recognition of key features’ is not directly described in the clinical reasoning literature as a theoretical entity in itself, the key features assessment is an accepted form of evaluation of clinical reasoning ability. It involves the ability to transform the presenting information into key clinical concepts, recognising what information is clinically significant and requires action. This emerged as significant from the analysis of the transcripts, as it appeared there were marked differences between the cohorts in their ability to do this. This is well illustrated by the differences in the recognition and management of the patient’s tachycardia.

The competent physician understands that fast AF is an important diagnosis that requires immediate recognition and management. In this virtual clinical case, the diagnosis of atrial fibrillation required the recognition of relevant information from the triage form, the patient history, clinical examination, and from appropriate investigations.

In the virtual case, the patient, an elderly woman, presented to the Emergency Department feeling generally unwell, with fatigue, occasional dyspnoea on mild exertion, a heart rate of approximately 135 beats per minute, and with an irregularly irregular pulse on examination. The case scenario dictates that the patient was confused and unable to provide a complete medical history. However, the virtual patient was able to recall her usual medications when asked (Betaloc® and aspirin), and these were also listed on her triage form. The patient was able to explain only that they were prescribed ‘for her heart’. She also volunteered freely in the history that she had forgotten to take her pills for a few days, because she was feeling unwell. An ECG recording was available if requested, and this showed atrial fibrillation, with a ventricular rate of approximately 135 beats per minute, (appendix 4).
All of the medically qualified participants immediately recognised that the virtual patient’s tachycardia required investigation and management.

Participants from the doctor cohort all recognised that her pre-existing rate-control and anticoagulant medications made AF a likely diagnosis. All of the doctors noted the clinical findings of tachycardia and an irregularly irregular pulse as consistent with AF. This group of participants also all ordered an ECG in to confirm atrial fibrillation as the cause of these observed clinical symptoms. All doctors recognised AF on the ECG. In addition, most of the medically qualified participants noted the significance of the patient’s recent non-adherence to her rate-control medication, and concluded that the observed high rate was a likely consequence of this. Most also commented that concurrent infection may have contributed to her rapid rate. Some of the doctor cohort also suggested that her fast atrial fibrillation might have had an ischaemic cause.

Some excerpts from the transcripts illustrate this below:

[Doctor 4] it sounds like she has also forgotten her medication (it sounds like she has AF, but I will double check with her GP) and her rate is currently 130, that could be from forgetting her meds, the infection or both.

This segment indicates that the participant recognised the medications as likely to have been prescribed for pre-existing AF, they note in context the fast heart rate, and show understanding why this is occurring in the clinical scenario.

This is reiterated below in another example from a transcript from a different participant:

[Doctor 2] Well she is in fast AF probably for 2 reasons: 1) she hasn't taken her betaloc for a couple of days and also she has this UTI. However I would want to exclude any ischaemic causes since she also had some SOB and her ECG may have some ST depression in V5 only
The following excerpt, (which was previously used as an example above to demonstrate elaborated knowledge), also illustrates the participant’s recognition of the significance of the medications.

[**House surgeon 11**] I’ve got a delightful 76 yr old lady. *She has known prev cardiac disease and is on aspirin and metoprolol for what sound slike AF*

This was sharply contrasted with the clinical behaviour of the fifth-year students, who rarely investigated the patient’s tachycardia, so generally no specific management of the AF ensued.

All of the fifth-year students recorded the virtual patient’s rate control and anticoagulation medications. Despite this, the majority of participants in this cohort did not diagnose AF explicitly, nor outline any treatment plans for the management of it. In addition most participants did not order an ECG in response to the tachycardia. There was the occasional diagnosis of AF made, one which was made from the history and clinical findings, but without ECG confirmation. Of the participants that did make the diagnosis of AF, it was generally not recognised that the concurrent infection may have contributed to its presentation.

Below are some excerpts from a transcript of one fifth-year participant that illustrates that while a participant accesses the relevant (although somewhat vague) information in the history, he either does not recognise its significance or disregards it when formulating diagnoses.

These were recorded during the initial part of the clinical case, while taking the patient’s history:

[**Gertrude Macfarlane**] *oh this whole thing started few days ago*

*i am now feeling more tired*

*haven’t been eating or drinking properly*
[Fifth-year student 1] Aha...

[Gertrude Macfarlane] i went bit short of breath yesterday while cleaning the garden

[Fifth-year student 1] that must be frightening

[Gertrude Macfarlane] yes, but it settled down when i sat down

This was continued 6 minutes later:

[Fifth-year student 1] ok, can you tell me after you sat down yesterday you felt better with your breathing. were you dizzy at all?

[Gertrude Macfarlane] no
i had a dizzy spell years ago
this wasn't like that

[Fifth-year student 1] alright, have you ever had any similar episodes of shortness of breath or going off your food and feeling tired?

[Gertrude Macfarlane] well i remember going short of breath when i had the dizzy spell, but no, i never had any eating problems
the doc gave me pills for the spell

7 minutes later this excerpt was recorded:

[Fifth-year student 1] now Mrs McFarlane you said you take pills for something but you dont know what it is for?

[Gertrude Macfarlane] yes
[Fifth-year student 1] ok do you know if you have any medical or surgical conditions?

[Gertrude Macfarlane] no
i take betaloc and cartia

While taking a history, this participant obtained some relevant information; including the patient’s current medications, a medical history suggestive of a chronic condition with recurrent dyspnoea, and current symptoms of fatigue and dyspnoea. After obtaining the history, and results of examinations (including taking the pulse and review of the ECG), the following excerpt was recorded:

[Fifth-year student 1] well I think she has got a UTI with secondary dehydration. this is because her urine dipstick is positive for WBC and she has got a Temp of 38.1. She is volume depleted and tachy.

Within the transcript, no further mention was made of the tachycardia. The admission note written by this participant includes these comments:
“PMHx: * she had pervious spells (for night house surgeon to chase up exact nature)” and “HR132 (h)”
However, no diagnosis of AF or management plan for the tachycardia was recorded at any stage. This suggests that this participant did not recognise the combination of findings as fast AF, or if so, did not think it important enough to investigate immediately.

Analysis of the transcripts showed that the third-year students also all recognised that the patient was tachycardic, but none were able to generate AF as a possible diagnosis for the tachycardia. Although they were unable to generate the diagnosis, most participants from this group did manage to elicit some relevant cardiac history from the patient.

This is demonstrated below.
[Third-year student 1] So, what were you doing when you first noticed the shortness of breath?

[Gertrude Macfarlane] I was collecting the dead leaves

[irrelevant chatter]

[Third-year student 1] So, what did you do at the point to reduce the SHORTNESS OF BREATH THAT YOU HAVE EXPERIENCED?

[Gertrude Macfarlane] Ooh nothing, I just sat down, and it went away quickly. Anyway I am taking this medicine as well. So I think it helps.

This was continued 4 minutes later:

[Third-year student 1] What was it like? Will you describe it for me?

[Gertrude Macfarlane] I missed breakfast as well. Ooh I don’t know how to describe. It was like panting you know [sic].

[Third-year student 1] I see.

[Gertrude Macfarlane] I had to heavily breath, and my heart waws bit funny going fast or slow.

[Third-year student 1] So, did you notice anything with the breathlessness?

[Gertrude Macfarlane] Heart rhythm was not full normal. I don’t emember much. Think I’m bit confused now.
[Third-year student 1] Was it beeting faster?

[general discussion about patient’s experience of confusion]

[Gertrude Macfarlane] yes, no breakfast, no medicine :(

[Third-year student 1] wha tkind of medicines you take?

[Gertrude Macfarlane] BETALOC and CARTIA
ONE OF EACH, ONCE A DAY

[Third-year student 1] Do you know what are they for? 
Is it diabetes?

[Gertrude Macfarlane] ooh my doctor gave them some tie ago, to control the
orregular heart beat
no, no diabetes

With this information, this participant made a diagnosis made 14 minutes later:

[Third-year student 1]: fem pt aged XX experienced SOB, brought by her neighbour
She had 2-3 episodes, on heart medic
Oh, she exp tachycardia with SOB when doing minmal exercise
I think she has angina!

This participant obviously elicited a good history of prior treatment of an arrhythmia, 
and yet diagnosed the tachycardia as angina. It seems as this diagnosis was made on 
the basis of the symptoms of dyspnoea, and recognising the medications as being a 
likely treatment for a condition with a cardiac basis.
The following excerpt illustrates a similar pattern from another third-year participant: This participant initially established within the history that the patient was experiencing dypsnoea.

[Third-year student 3] Have you noticed any other changes in the last few days?

[Gertrude Macfarlane] Umm...
Well I was huffing and puffing a bit when I got to the chair here
But I caught my breath very quickly
I don’t know if that’s important?

[Third-year student 3] Ok is that the first time you have felt like you were huffing and puffing like this?

[Gertrude Macfarlane] Oh maybe yesterday as well, just when I went down to the mailbox
There wasn’t any mail, I don’t think, but I was a little breathless then as well

[Third-year student 3] Ok and did you notice any chest pain?

[Gertrude Macfarlane] No, oh no I’ve never had that. Not at least that I know of

[Third-year student 3] Ok that’s fine. I’ll ask you some general questions about different body systems if that’s ok with you ;)

Further information relevant to the patient’s cardiac history was obtained 15 minutes later:

[Third-year student 3] Do you have any medical conditions?

[Gertrude Macfarlane] No I don’t think so
[Third-year student 3] Ok and are you taking any medication?

[Gertrude Macfarlane] Oh yes
I’m taking aspirin for my heart
And Betaloc for the rhythm, I think

[Third-year student 3] Ok is that all?

[Gertrude Macfarlane] Umm
Sometimes panadol for a headache
But not today

[Third-year student 3] Ok. How long have you been taking the aspirin and Betaloc for?

[Gertrude Macfarlane] Oh years

From this information, as well as examination findings, the following diagnoses were made:

[Registrar] what do you think is going on with mrs mcf

[Third-year student 3] I think she might have a UTI

[Registrar] okay.. can you tell me why you think that?

[Third-year student 3] I found that she has been feeling tired, confused and generally unwell for the last few days. she has a temperature, is tachycardic and has a positive urine dipstick for both protein and leucocytes. she has increased her urination frequency, although the volumes have decreased. she also expressed the urine had a foul smell. Chest and Heart both appear normal.
[Registrar] ok that all sounds good. Do you think she has any other problems? or do you have any differential diagnoses

[Third-year student 3] I thought perhaps dementia, diabetes mellitus, a thyroid issue or alzheimers, although her findings generally suggest a UTI.

It would appear that this participant had obtained the relevant information; the patient was experiencing dyspnoea, she had been prescribed Betaloc® for ‘rhythm’, and she was noted to be currently tachycardic. From this, the participant concluded that ‘chest and heart appear normal’, and that the tachycardia was potentially due to ‘thyroid issues’.

Other participants from this cohort also recognised the patient’s tachycardia, but none suggested AF as a possible diagnosis. Other diagnoses proposed included:

[Third-year student 4] hypovolemia

[Third-year student 3] potentially thyroid issues due to her tiredness and tachycardia.

One participant in this cohort suggested the observed tachycardia was due to cessation of the rate control medication, but still did not specify that the patient was in AF.

[Third-year student 4] tachycardia may be due to discontinuation of beta blocker or due to dehydration.

Most participants in this group did not order an ECG in response to the tachycardia. If an ECG was obtained, those participants were not able to identify the rhythm as AF from the reading.
While recognition and management of AF can be analysed as a key feature of this case, it also illustrates development of an illness script. The doctor cohort identifies the presenting information of ‘tachycardic’ and ‘concurrent illness’ and ‘non-adherence to rate-control medications’ as being likely fast AF.

In summary, while the majority of participants from all groups elicited the same information in the history and examination, only the medically experienced participants were able to abstract the clinical importance of the tachycardia, and the necessity to act on these findings.
3.2.2 Thematic analysis of admission/discharge notes

The final endpoint of the clinical case was the admission or discharge of the virtual patient. To complete this, all participants were required to provide a written summary note. The summaries provided a concise written insight into the clinical thinking of the participants, and a rapid means of assessing their clinical competence. Analysis of the summary notes showed marked differences between those written by medical students compared with those produced by experienced doctors.

A brief qualitative analysis of the notes was undertaken to delineate the differences, again using a general inductive approach to establish key themes and patterns.

Key areas of difference apparent between the cohorts were in the overall structure of the notes, the content of the presenting problem section, the impressions or diagnoses offered, and in the management plans. These are described in detail below, and interpretation of these differences with respect to clinical reasoning ability is further examined in the discussion.

3.2.2.1 Overall structure of admission notes

Most medical admission notes are clearly structured in a universally recognised fashion. It is generally expected that they contain, as a minimum, demographic data of the patient, presenting problems, relevant background medical history, medications, allergies and social history. There is usually a concise summary of examination findings and investigations ordered, with differential diagnoses or impressions, followed by a management plan.

This part of the analysis reviews only the structure of the notes; the order and layout of information and types of information included (e.g. past medical history,
examination findings). Further analysis of the content of the subsections is discussed later.

The participants from the doctor group all produced admission notes with a clear structure, containing all of the relevant sections and information. Fifth-year students were also usually able to structure their admission notes logically, and included most of the relevant parts. Third-year students were unable to write a clearly structured admission note, and omitted key sections of information.

Supporting evidence from the summary notes:

**Doctors:**
All admission notes from this group of participants contained a clearly structured *history of presenting complaint, past medical history and medications*, and the note generally also included the social setting and effect on the patient as relevant information.

All notes from this group included clearly titled investigation and examination findings, and all offered impressions that encompassed all of the presenting problems. The majority of admission notes in this group concluded with separate, well-structured management plans.

**Fifth-year students:**
All notes produced by this cohort appeared to have a logical structure, with patient history, examination and investigations more often presented separately, (although not always labelled). The history section of the note was sometimes split up into its individual components such as *presenting complaint, background medical history* and *medication*. Some participants from this group also included impressions, with most also incorporating some kind of management plan.

**Third-year Students:**
Students from this group were unable to produce a clearly structured admission note containing all relevant information.

While most participants in this group started their admission notes with the patient’s presenting symptoms, there was otherwise no overall clear structure. Some of the participants included medical history or medications, some examination and investigation findings, and there were occasional attempts at cursory plans. Often all of these aspects were mixed together without any headings or separation, and none of the third-year student notes included all parts expected to be present in an admission note. Most of the notes contained a single diagnosis, but no problem lists or impressions were offered.

3.2.2.2 Structure and content of the ‘presenting problem’

The presenting problem (also known as the History of Presenting Complaint (HPC), or Presenting Complaint (PC)) usually contains a succinct summary of aspects of the patient history grouped into recognisable clinical entities. It would usually include an explanation of key presenting symptoms, their duration and sequence of events, relevant background medical history and note of any important symptom absences (relevant negatives).

The medically qualified participants were able to summarise and communicate most relevant aspects of the presenting problem. Again, the fifth-year medical students were also able to include most salient points within the presenting complaint. However the third-year medical students were unable to communicate a concise summary of the patient’s presenting problems:

Supporting evidence from the summary notes:

Doctors:
The presenting problem section of an admission note written by a medically qualified participant is shown below:

‘76 y.o. female with background of fast AF on Betaloc, no other medical conditions. PC: ~5/7 unwell, off food/drink and ?confusion per neighbour (forgot to feed her cat). Also admits to foul-smelling urine. Briefly dyspnoeic while gardening yesterday. No chest pain, no cough, no abdo symptoms. Thirsty. Missed her aspirin and Betaloc this morning.’

All admissions written by the participants in this group began with a clearly structured and titled *History of Presenting Complaint*.

Within this section, all stated clearly the patient age, duration of illness and the key presenting symptoms (confusion and/or generally unwell). All participants from this group included as relevant that the patient was not eating and drinking, and had not been taking her medications. The majority also noted the urinary symptoms and dyspnoea, and some included (within the presenting problem) the background of AF and relevant negatives from the history.

Doctors are also better able to communicate a much larger amount of relevant information quickly. This is illustrated by another example from an admission note, which includes all relevant information in less than 100 words.

‘76 yr old lady. Found confused by neighbour and brought into ED. Reports 2-3 day history of feeling generally unwell, hot, short of breath on exertion, confused. Off her food and not drinking. Has noted some palpitations, and short of breath going to mail box, not usual for her. Denies chest pain, ankle swelling, cough or sputum. Has forgotten to take her medication for last few days. Reports funny feeling on passing urine and odd smell, no frequency, passing less than usual, small amounts at a time. Denies abdominal pain, change in bowel, vomiting, slightly nauseated.’
Fifth-year students:

‘Mrs McFarlane is a 79 year old female that presented to ED with decreased energy, shortness of breath and confusion. This started 3 days ago and has been associated with increased thirst, decreased urine output and also discomfort on passing urine.’

All admission notes from this group started with presenting problem but only some had it clearly structured separately from the rest of the history, and only a single participant had it labelled.

Within what appeared to be the presenting problem section, all notes from this group contained the basic information of patient age and the duration of the current illness. They also all outlined the main presenting symptoms as confusion or fatigue or being generally unwell, and all also mentioned the urinary symptoms and dyspnoea.

Some participants noted the patient’s inability to care for herself, and there were isolated comments about the history of AF or relevant negatives from other systems.

No participants from this group mentioned non-adherence to medication as relevant to the presenting problem.

Third-year Students:

This is an example of what appeared to be the ‘presenting problem’ section of an admission note from a third-year participant:

‘Patient have been feeling sleepy for the past 2-3 days due to tiredness. She sleeps in the morning and went to bed quite early. Neighbour explained that she looks confused and ask patient to go see the doctor. Other than that she was feeling breathless with occasional palpitations on mild exertion which goes pass on rest. She sleeps with 2 pillows (normal for her). She was also complained of feeling thirsty
than usual. When asked about her urine, she haven’t peed a lot and that her urine smells “funny.”

Patient did know the cause of her symptoms and that there was no events prior to that could explain her symptoms.’

This part of the admission note was over 100 words, and contains far less relevant information than that produced by participants from the doctor cohort.

As discussed above, there was minimal structure to the admission or discharge notes from this cohort. While all of the notes started with an attempt at a summary of the presenting symptoms, they were often in a narrative style rather than a concise presentation of the key issues. There was no real organisation of the presenting symptoms into recognisable clinical concepts.

Only one participant attempted a structured separate presenting complaint, but it contained a mixture of symptoms, signs and examination findings, rather than organisation of relevant parts of the patient history as is usually expected.

All of the third-year participants noted the presenting problems as ‘feeling unwell, confusion and/or fatigue’, and stated this somewhere within the section that appeared to be the history of presenting complaint. All participants in this group also noted dyspnoea as a key symptom.

The majority of participants in this group also included demographic details such as the patient age (some incorrectly recorded), and stated the duration of illness.

Only some of the third-year students mentioned urinary symptoms as relevant to the presenting problems, and there was an occasional note of the patient’s inability to care for herself or non-adherence to medications.
No participants from the third-year student group noted (as relevant to the presenting complaint), the patient’s history of AF, or stated the absence of relevant cardiovascular, respiratory or abdominal symptoms.

3.2.2.3 Diagnoses/impressions

An admission or discharge note is expected to contain a diagnosis, or as a minimum, an impression, (preliminary diagnosis). Sometimes this is preceded by a ‘problem list’, which contains a list of all the current medical issues of the patient. It is also sometimes acceptable to combine the problem list with the impressions.

Qualified doctors all provide detailed and explanatory impressions. Fifth-year students propose brief diagnoses that are not elaborated, and third-year students offer only a single diagnosis.

Supporting evidence from the summary notes:

Doctors:

All participants in this group supplied impressions that included diagnoses that could explain all of the presenting problems, in a unified way. Impressions from this cohort were generally well detailed, and some were structured with problems numbered and ordered.

Impressions often included pathophysiological links, such as ‘AF – likely long standing, fast because ?forgot medication, infection or both’. Impressions also sometimes incorporated current status of treatment. This is an example of the impression recorded by a medical participant:

‘Imp:

1) UTI

-started on IV Augmentin 1.2g Q8h
- fluid resus

2) Fast AF

-Prob secondary to UTI and not taking meds
- started on short acting metoprolol 25mg TDS
- Loaded with Aspirin (CHADSII scoer: 1)
- TnT pending to exclude ischaemic causes’

Fifth-year students:

Only some participants in this group actually stated any diagnoses explicitly.
Of the diagnoses available, they were generally quite brief, such as ‘Impression: ?infection ?renal pathology’.

Only one participant offered an impression encompassing all symptoms, or attempted to link the presenting issues: ‘Imp: Infective process (UTI) leading to confusion. Dehydration after minimal food/water, with subsequent AF.’

Third-year students:

Participants from this cohort mostly offered only a single diagnosis such as ‘urinary tract infection’, or occasionally attempted at a very basic level to unify the presenting problems, i.e. ‘hypovolaemia secondary to infection’. The diagnosis was usually not stated as a separate entity, but part of a narrative within the overall note.
This is an example of how a diagnosis appears within an admission note written by a third-year student:

‘Her thirst has been accompanied by an increased frequency in urination- with a decreased frequency in volume. She noticed the urine also had a foul smell. A urine dipstick was positive for leucocytes and protein – suggesting a UTI.’
3.2.2.4 Management plans

The final section of an admission note is the proposed plan of treatment for the patient. This is to communicate the intended management to other attending professionals, as well as a clear outline of any issues that may need further investigation or attention. It generally will also summarise any treatment already given.

Analysis of the management plan section showed a marked difference between those produced by the qualified doctors compared with the student cohorts. Doctors were able to generally provide clear and specific management plans, while those produced by fifth-year students were much broader, omitting key clinical issues. Third-year medical students were unable to formulate any form of adequate management plan.

Supporting evidence from the summary notes:

**Doctors:**

Most participants from this group offered clear and specific management plans that generally included all aspects of care required for the patient. This included documentation of both immediate care (fluids, antibiotics) and continuing care (follow up of the result of the urine culture, and some form of cardiac monitoring such as repeat ECGs or repeated Troponin T blood tests). Often a discussion of long-term care was also included: some noted the need for anticoagulation and on-going rate control for the patient’s AF. Occasional directions for nursing staff were included by participants from this group.

This is an example of a management plan formulated by a medically qualified participant:

`Plan`
1. Admit
2. IV fluids
3. Await urine culture
4. Cipro
5. Oral metoprolol and aspirin
6. Consider IV metoprolol if rate doesn’t improve with orals and fluid
7. Repeat ECG 30 min
8. Monitor obs closely
9. Consider warfarin

**Fifth-year Students:**

Most admission notes from this group included some kind of structured management plan, and these plans included immediate treatment such as intravenous fluids and antibiotics. Some participants noted the need for continuing care such as ‘chase urine culture’ and ‘await sensitivities’.

The management plans from this group still contained some quite general directions such as ‘monitoring’ or ‘observation’, and ‘review with consultant’ or ‘management of AF’, rather than a specific plan of action. All management plans from this cohort still had key clinical omissions, namely no clear management of the patient’s tachycardia, or plan for specific rate control, however this was possibly due to the poor diagnosis of the AF initially. Below is an example of a management plan written by a fifth-year participant:

‘Plan:
Admit to ward
give IV fluids 500ml and encourage oral intake
start on IV augmentin as per protocol awaiting sensitivities
Review with consultant in the morning.’

**Third-year Students:**
The summary notes from this group did not include any management plans. There was the occasional suggestion of ‘monitoring’ as the reason for admission. Below is an excerpt from an admission note that was possibly an attempt at a management plan by a participant from the third-year student group:

‘The reason for admission is because she is confused and tired and it would be unsafe to send her home at this stage. It would be better to have her monitored and if she feels better then send her home.’

3.2.2.5 Summary of differences between cohorts

In summary, the admission (or discharge) note produced at the end of the case provided a succinct insight into many aspects of the participant’s clinical competence. With increasing experience, the notes became more clearly structured, contained more relevant information, especially information pertaining to description of the presenting problems. The competent participant listed well-elaborated diagnoses with clearly outlined management plans. Many of the features from the quantitative score-sheet could be assessed quickly from the summary note, and aspects of clinical reasoning could also be inferred. While the pathway to diagnoses is not always evident, the accuracy of the diagnoses and identification of key features is usually apparent. It would appear that the summary notes represent most of the clinical thinking of the participant, and provide a quicker means of analysis.
Chapter Four: Discussion

4.1 Overview

This study assessed the validity of a virtual reality based clinical case as a form of assessment of clinical competence. In addition, this study specifically examined the capacity of the virtual reality case format to assess clinical reasoning ability.

Current modes of assessment are acknowledged to be limited in their ability to assess many aspects of clinical competence. The optimal method of assessment of clinical reasoning ability is still contentious. Also important is the development of tools to assess clinical decision-making ability, triaging, tolerance of ambiguity, and the ability to communicate succinctly and with clarity to colleagues.

The main findings from this study suggest that virtual reality based clinical cases may provide a novel and valid means to assess aspects of clinical competence. The virtual case-based assessment encompasses more aspects of competence than traditional means of assessment. In addition the virtual cases are able to provide unique data for the evaluation of participants' clinical reasoning abilities.

Results from this study showed that participants exhibited significant improvements in quantitatively assessed clinical performance with increasing medical training. Qualitative analysis of clinical reasoning themes indicated that with increasing medical experience, there was evidence of improved clinical reasoning ability - participants were better able to transform information into key clinical concepts, generate more accurate diagnoses, and generate correct diagnoses more efficiently.

However, the generalisability of the findings should be interpreted with caution in view of the small scope of the project, which was limited by the time taken for testing. With twelve participants in total, the project was intended only to be a pilot
study. With the small numbers of participants, obviously conclusions drawn from the results will be limited in power. While discussed in detail below, it is acknowledged that the validity of these results is limited by both small numbers of participants, and by case-specificity.

Despite these limitations, the initial results of this pilot study appear promising and suggest further research in this area is warranted.

This discussion intends to examine in detail whether these findings do provide acceptable evidence of the validity of the tool as an assessment method. This includes whether validity was established by comparison of the cohorts at different levels of medical training, and whether the cohorts were truly representative of different levels of expertise. Also examined is the use of a quantitative scorecard to measure performance, and evidence for construct validity of the assessment is evaluated.

The other key issue to be discussed is the usefulness of the tool to assess clinical reasoning ability. The supporting findings from the study are examined with reference to the theoretical clinical reasoning literature.

The practicalities of implementing virtual cases as an assessment tool are then considered. There is discussion of how sub-scores and analysis of summary documents may be employed as more efficient means of assessment than scorecards and transcript analysis. Finally, the main limitations of the study are outlined, and suggestions for further work are proposed.

4.2 Evidence of validity

The virtual reality based case appeared to have high face validity, as the simulation was a close approximation to management of a real clinical scenario, performed in real time.
The construct validity of the assessment was inferred from the ability of the experts (the qualified doctor cohort) to perform better than the novice student cohorts. This was primarily ascertained by scoring the achievement of set outcomes. Contrasting the performance of cohorts is an accepted method of establishing construct validity, especially when there is no other form of assessment that can be reasonably be used as a comparison. This method of establishing validity was chosen because previous research indicated other forms of validation have been problematic. Establishing concurrent validity by comparison to other assessments is difficult, primarily because the virtual patient format is unique and assesses components of competence that cannot yet be tested by any other means.

It would appear that the assumption that validity could be established by comparing the performance of cohorts of different levels of training was reasonable. However, two issues arose from this assumption. It was necessary to first establish that the cohorts were representative of different stages of development of competence. Relevant to this project, a higher proportion of the third-year student cohort were international students, with English not being their first language. While the medical course is taught only in English, it may be that in the early stages, language barriers may have influenced the overall performance of participants in this cohort, hence overestimating the magnitude of the difference observed. Participants in the doctor cohort varied considerably in their clinical experience, but as they all had attained general scope with the Medical Council of New Zealand, it appeared reasonable to assume this represented sufficient competence to be deemed experts, as the assessment was benchmarked at the level of a junior doctor.

The other important issue was to establish whether the virtual case tested the correct construct, or whether it reflected some other difference between the cohorts such as their ability to use the technology. As performance was seen to improve with increasing clinical experience, it appears unlikely that this was due to the more junior cohort being less able to operate the VR technology. In fact, it is
likely that the younger student cohorts would be more familiar with the use of the technology, which may have conferred some advantage.

It is worth noting that the process of testing itself has an effect on performance; it may either underestimate ability due to performance anxiety, or sometimes the attention necessary to complete the simulation has a positive effect on performance. However, in this project, there does not seem to be any reason to believe that the testing process would affect cohorts differently.

4.3 Validity of the scorecard as a tool for assessment of clinical performance

Quantitative scoring of overall performance in the clinical case was implemented using the scorecard developed from expert opinion prior to testing. Results from the quantitative scoring clearly differentiated between the performance of third-year medical students and that of the qualified doctors, with the doctor group achieving the highest average scores. Fifth-year students produced an average score that appeared to lie between those from the other cohorts, but numbers were too small for this difference to attain statistical significance.

Overall these results suggest that the format of the assessment has construct validity. There is clear demonstration that experts significantly out-perform novices. Naturally it is expected that qualified doctors will exhibit more clinical competence than junior medical students, so these results support the hypothesis that this form of evaluation does indeed measure a difference in clinical competence.

There have been concerns about the validity and usefulness of attempting to assess competence using quantitative scoring systems in simulations. Previous forms of scorecard-type assessment designed to measure diagnostic ability have previously been developed and then abandoned. The principal problems with using scoring systems for assessment were that they required a consensus from experts about the optimal strategy through a given problem, and often involved complex weighting
schemes. As discussed in the introduction, scoring in this manner was thought to award thoroughness and penalise efficiency.\(^67\)

To avoid this issue, this case was specifically designed to be a poorly differentiated clinical problem, so thoroughness was necessary to resolve the clinical problem. Also built into this scorecard were penalties for unnecessary investigations or treatments. One crucial difference with this form of assessment compared with the scoring of the original written simulations was that patient management occurred in real-time, so efficiency could be noted by measuring the time taken to complete the clinical case. In addition, the scoring system also included aspects such as inclusion of differential diagnoses and management decisions. This was intended to provide a better evaluation of actual clinical practice, rather than to be a system that assessed diagnostic thinking in isolation.

Consistent with the literature suggesting that clinical problem solving occurs in an idiosyncratic fashion,\(^19\) there was no assessment of the pathway through the problem, only of expected outcomes. However, these outcomes did include eliciting most aspects of the medical history, and the main examination and investigations: these were necessary to make the diagnoses and exclude the main differential diagnoses.

In some ways, the scorecard used in this project was a hybrid of a standard scoring system with a key features type assessment,\(^22,28\) in an attempt to utilise the most useful aspects of both schemes. It was not a true key features type problem as described in the literature, as typically the key features format would only focus on critical aspects of the case, generally a small number of items.\(^22,28\) While still focussing on outcomes that were considered necessary to complete the case competently, the scoring system developed for testing in this project evaluated the achievement of a considerably greater number of outcomes, including features that were perhaps not critical but still desirable. Another difference was in the delivery of initial information; key features problems are generally presented in a written or computerised format, and require a SAQ format in an attempt to avoid cueing.\(^22\)
This format of this virtual clinical case allowed examinees to freely elicit information in an open-ended manner, with the history being entirely directed by the participant, minimising the problem of cueing. It is possible however that some cueing may have occurred in the examination phase of the simulation, as only relevant examinations were offered within the programme.

Development of a scoring system for this project was in keeping with recommendations from the most recent consensus statement and recommendations from the Ottawa Conference (2010) for technology-enabled assessment of health professions education: ‘Researchers should develop scoring methods that automate the collection, integration and analysis of the vast and often novel information available through technology enabled assessment.’

The overall results from this project seem to reflect expected differences in competence; hence this may mean that for this particular clinical case, the scorecard is a more comprehensive and robust tool than any previously developed. The findings corroborate the existing literature that suggests virtual patients are likely to offer an effective means by which clinical competence can be assessed. This study shows some evidence of validity of the virtual case tool, which contrasts with the early work in the area, which failed to demonstrate that virtual patients were a valid form of assessment of competence.

It is still possible that the overall score purely reflects the development of underlying knowledge rather than overall clinical competence or reasoning ability, as it would be expected that the knowledge base would also increase with level of training. However the thematic analysis indicates that clinical reasoning ability was observed to improve with clinical experience. This is discussed further below.

It has been previously observed that medical students at graduating level have been unable to resolve an atypical problem with presented in the key feature format. The results from this study perhaps reflect these findings: this particular case was not atypical but rather was not clearly defined, and the majority of tested students
were unable to complete it satisfactorily. This may not necessarily reflect a problem with the format of the testing, but perhaps indicates that graduating students have not yet developed the clinical reasoning ability to deal with atypical or poorly defined clinical problems.

4.4 Specific assessment of clinical reasoning ability:

4.4.1 Quantitative measures of clinical reasoning

While not assessed formally in a quantitative manner, some outcomes consistent with clinical reasoning ability were evident from the scorecard used in the virtual case. Within the scorecard was a section for separate assessment of achievement of the correct diagnoses, and points were awarded for supplying adequate differential diagnoses. Depending on the definition of clinical reasoning, (often it is thought to also encompass the cognitive planning associated with treatment as well as diagnosis), scoring of achieving effective management could also be assessed. While these aspects may be useful in providing an overview of clinical reasoning ability, it is likely to be better assessed by the qualitative methods.

4.4.2 Measures of efficiency by time taken

The only quantitative measure recorded that could reflect clinical reasoning ability was the measure of times taken to make diagnoses and complete the case. However, in this study, no significant differences were found between cohorts in the times taken to achieve these outcomes, most likely due to the small numbers of participants tested. It is entirely possible that assessment of time taken was not useful in this setting, as it was primarily confounded by the keyboarding ability of the participants, which varied considerably, and in no particular relationship to level of training.
Of note, while results did not show the differences to be significant, it appeared that the doctors, on average, were 20 minutes faster to make the diagnoses than the student groups, even when compared with the student groups who made incorrect and incomplete diagnoses. This apparent trend would have possibly been confirmed if larger numbers had been tested.

While again not significant, results suggest that the time taken to complete the case by the different cohorts could be graphically represented by an inverted U-type relationship. The third-year medical students completed the case slightly more quickly than the senior students, and it is theorised that this happened because the junior medical students omitted large chunks of information due to their lack of knowledge, or because they neglected many of the clinical management steps. The fifth-year students generally appeared to take the longest time; this may have occurred because this group had now attained more knowledge but insufficient experience to access it quickly. The doctor cohort completed the case with the fastest average time, possibly exhibiting efficiency related to their non-analytical reasoning skills, and experience in management of similar cases.

It is difficult to draw any conclusions about the time taken to complete the case if achievement of safe effective management is taken into account. None of the third-year students and only one fifth-year student managed to meet the clinical objectives, so it is not possible to assess the time taken to do so in any meaningful way.

4.4.3 Assessment of clinical reasoning skills by analysis of transcripts

Using a virtual reality based clinical case to assess clinical reasoning ability is a novel concept. This form of assessment supplies information about many aspects of competence that is usually difficult to assess. Using virtual reality to represent a clinical case is essentially an enhancement of the virtual patient format, already acknowledged for its usefulness for evaluating clinical reasoning ability.21,39
Like the well-described virtual patient, one virtual reality based case provides a huge amount of data about the participant’s clinical thinking and behaviour. With each case generating on average, over 3000 words of transcript, it was possible to analyse the extent and sequence of clinical history taking, examination and investigations undertaken, as well as prioritising of management decisions. Not only that, but it allowed the examiner to view the overall pathway taken through a problem. It could be observed what information was available and processed before decisions were made, and allowed the examiner to specifically ask for justification and elaboration for those decisions. As the live patient allowed non-automated history taking, offering only the information that was requested, this completely avoided cueing. The entire case was open-ended, entirely directed by the participant’s own decisions. This means that the clinical behaviour was likely to be much more authentic, much closer to Miller’s ‘does’ part of the pyramid.18

The findings from this study were consistent with current theories of clinical reasoning. With increasing competence, participants demonstrated more proficiency in non-analytical reasoning skills. The experts exhibited a more direct pathway to diagnosis, illustrating the likely underlying formation of illness scripts. The experts were able to elaborate their knowledge more clearly to make more accurate, unified and well-supported diagnoses, and they could clearly identify the key features of the case, resulting in better patient outcomes.

The format of this assessment provided some unique insights into the participants’ clinical thinking. The findings of this study suggest that the part of the construct assessed was actually clinical reasoning. Changes in clinical behaviour were observed with increasing experience. These observed differences in clinical thinking were in keeping with what would be predicted by current theories describing the development of clinical reasoning ability.

This is consistent with current literature that suggests simulated virtual patient scenarios (including those utilising virtual reality) are likely to promote clinical reasoning ability. There is little published evidence yet supporting the use of virtual
patients to assess clinical reasoning ability, although it has been suggested to be, in concept at least, theoretically sound.²¹

It seems then, that by using a virtual reality format, this study has examined a completely novel method of assessing clinical reasoning. Potentially this format may offer the most integrative type of assessment yet available, as it assesses not only diagnostic aptitude, but also the ability to elicit relevant information, and clinical decision-making in real time. The format of the simulation is similar to that used by Imperial College in London⁵⁷ (patient-focussed simulation), but is unique in that it has been utilised as an assessment tool.

4.5 Practicalities of implementation

One major disadvantage of this virtual case format is the time required for analysis. While a non-expert would most likely be capable of scoring participants using the Clinical Performance Scorecard, extracting results from the transcripts is time intensive. However there may be a number of feasible alternative options for faster analysis that would enhance the practicality of using a virtual case as an assessment tool. These include using only sub-scores that reflect just the diagnostic and management decisions, and analysis using only the summary note (admission or discharge note) produced at the completion of the clinical case.

4.5.1 The value of sub-scores: inferences from the diagnostic and management decisions

There was a significant difference between the cohorts in their ability to generate full and correct diagnoses in the virtual case. As expected, doctors outperformed students, reflecting again a difference in competence with stage of training. Significantly, achievement of safe effective management was only achieved by participants from the cohort of doctors, with the sole exception of one fifth-year student. These results seem to reflect a crucial difference between the abilities of
students compared with doctors. It is most likely due to the effect of clinical experience, and this possibly represents a good endpoint measurement for high-stakes exit exams, as it seems to discriminate most clearly the clinical competence expected at the level of a graduating medical student.

4.5.2 The summary document: A useful insight into overall performance

Review of the extensive case transcripts was time-consuming. However, each participant was required to complete the case by submitting a summary note in the form of an admission or discharge note. Analysis of these summary documents indicated that they contained the majority of relevant information needed to evaluate performance, without the need to analyse the transcripts in full.

Analysis of the summary documents indicated that clinical competence was rapidly evident from the structure and content of the admission (or discharge) notes. Results indicated that clinical competence could be inferred from the production of a clearly structured note that was initiated with a succinct description of the presenting problems. The summary note produced by the competent practitioner also included all relevant issues from the history, examination and investigations, followed by a clear overall impression or diagnosis, usually accompanied by a list of logical differential diagnoses. A detailed and safe management plan would provide additional evidence of clinical competence.

It did not appear that it was merely the ability to communicate medical information concisely that altered with training. Evidence of clinical reasoning ability emerged with the production of elaborated diagnoses, and clear identification of key features – these were often apparent in the diagnoses and documentation of clinical decisions.

Given the findings of this analysis it is conceivable that a form of assessment based only on the summary note could be developed. This would have advantages in
terms of time efficiency for the examiners. Criteria for assessment could include inclusion of key aspects from the history, examination and investigations, clear and accurate differential diagnoses, and a coherent logical and safe management plan. If this method was employed, there could also be the option of reviewing the complete transcripts for only those participants whose performance was borderline.

4.6 Limitations to study

The key limitation to this study was the case specificity of the assessment format. As discussed in the literature, the results only reflect competence specific to ability with one scenario, so it is not possible to generalise without testing across multiple cases. The other main limitation was the small number of participants.

Other issues include possible problems with reliability due to differences between the two different actors, and by potential differences in the way the patient was portrayed to each subject. While a pre-scripted scenario was used, due to the open questioning format, there were inevitably differences in the way the standardised patient presented their issues on each occasion. It was also possible that the blinding of the actor to the participants’ identities was jeopardised at times. One of the standardised patients, who is an experienced actor, was also a member of the medical teaching staff, and potentially could have recognized some of the participants from their responses.

Some participants experienced minor difficulties in navigating within the virtual hospital, and required some assistance to understand what was available within the clinical case. Examples of this included participants who attempted intravenous fluid resuscitation, but did not first insert a cannula with which to administer the fluids. This was likely due to the virtual nature of the case, as it was not immediately apparent that cannulation was necessary in a virtual world. Other participants may have neglected to order investigations such as ECGs; this was because the ECG machine was not immediately apparent when interacting with the patient. This
provided a potential bias: the more experienced participants were aware that an ECG machine should be available in an emergency department, so made an effort to locate it, but this may not have occurred to medical students with no clinical experience.

As already discussed, some participants, primarily from the third-year student cohort, had English as a second language. This may have adversely affected their performance, although probably not to a large extent since they were selected from an English speaking medical school.

There was a predominance of male participants in this study. This was thought to be a selection bias. The study recruited volunteers, so tended to select those more naturally inclined towards virtual reality and gaming, which is often assumed to be a masculine trait. However, increasing proportions of ‘gamers’ are now female, so this assumption may be incorrect. Another consequence of using volunteers is the possibility of selection bias towards those more likely to have aptitude with virtual reality programmes, however this should not have affected the comparison between cohorts, but may have implications for generalisability of the tool.

There were some technical difficulties experienced by the participants while using the virtual case. Some subjects experienced difficulties in reading the ECG and triage notes due to the poor resolution available within the programme. Occasionally there were also network issues, resulting in either the patient or the participant losing the online connection, necessitating the case to be paused until it could be resumed. This may have influenced the thinking processes of the participant, however it happened rarely.

While most participants expressed enthusiasm and approval for the format of the virtual case, there were some complaints about the need to take the medical history by text. Most participants stated that it took much longer than a real life history, and they were unable to pick up visual cues from the patient, or to do other tasks such as examination concurrently. These were technological limitations to the
programme. Other technological limitations meant that the patient was not able to respond visibly when given treatment such as oxygen. It is acknowledged that these are issues that should be addressed, and with technological advances may well be overcome.

4.7 Future directions

Further evidence of validity of the virtual reality based clinical cases, as a means of assessment of clinical competence is needed. To provide this, multiple cases should be tested, with a greater number of participants.

To enhance the experience of the virtual case, some technological improvements would be desirable. As simulation technology improves, it may be possible to replace the text-based history-taking with a real-time audible history. It may in the future be possible for the simulated patient to provide a spoken history in response to questioning, identical to a real clinical situation. The actor could have their own features overlaid by the simulated face of the patient, (so-called ‘augmented reality’). It is highly probable that advances in technology will soon allow much higher fidelity of simulation, complete with virtual patients in high definition exhibiting expressions, and signs of the disease as required. For example, the virtual patient suffering a myocardial infarction would appear pale and diaphoretic, and would show evidence of clinical improvement as the oxygen and morphine were administered. The ability to take an oral history would also enable the assessments to be completed in a shorter time.

4.8 Conclusion

Although the current project was only a preliminary study, the virtual reality based clinical case is a promising new format for standardised assessment of clinical competence. It provides a completely novel approach to the evaluation of clinical reasoning ability. There is modest evidence that the virtual case is a valid form of
assessment of clinical competence, and may provide an innovative means of assessment of a participant’s clinical reasoning processes. This mode of testing encompasses many aspects of clinical competence that current methods are unable to measure, such as clinical-decision making in real-time. As stated in the introduction: ‘Assessment at the apex of Miller’s pyramid, the ‘does’, is the international challenge of the century for all involved in clinical competence-testing.’ This pilot study may be one of the first steps towards meeting that challenge.
Bibliography


Dear Professor Walker,

I am writing to let you know that, at its recent meeting, the Ethics Committee considered your proposal entitled “Performance of medical students and doctors in virtual reality simulated cases: A pilot study for developing a tool for the assessment of clinical competence”.

As a result of that consideration, the current status of your proposal is:- Approved

For your future reference, the Ethics Committee’s reference code for this project is:- 12/039.

The comments and views expressed by the Ethics Committee concerning your proposal are as follows:-

While approving the application, the Committee would be grateful if you would respond to the following:

The Committee is interested as to how the results of the research will be applied. How will the results of the virtual cases be compared or adapted to a real life setting?

Approval is for up to three years from the date of this letter. If this project has not been completed within three years from the date of this letter, re-approval must be requested. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing.
Yours sincerely,

Mr Gary Witte  
**Manager, Academic Committees**  
Tel: 479 8256  
Email: gary.witte@otago.ac.nz

c.c. Dr J B Adams  Dean  Dunedin School of Medicine
Appendix 2.
Otago Virtual Hospital scenario for actors playing Gertrude MacFarlane:

YOU are the only person that gets to read this... Hi and welcome to the world of Gertrude MacFarlane. This notecard contains important parts of the back story for Gertrude. The parts in CAPS are the important information that you must share accurately with your colleagues. The paperwork, lab values, x-rays, etc depend on these presentation aspects. How you play the rest of Gertrude is over to you. So, while Gertrude presents with some confusion, how much confusion you present with is over to you. You are basically a fit and well older lady. You have long standing Atrial Fibrillation, have developed a UTI and become generally unwell over the last few days. Your unwellness is showing as being: more TIRED than usual (you had a snooze the other afternoon and you don’t usually do this, you fell asleep in front of the TV last night, and have slept in on each of the last couple of mornings); a little SHORT OF BREATH ON EXERTION, but this SETTLES QUICKLY when you rest (because you’ve forgotten your pills you’re in a faster AFib and get SOB with exercise); and you’re just not feeling quite your usual self.

On questioning you will admit to: NOT EATING OR DRINKING AS MUCH AS USUAL (“I just don’t feel like it, ... but right now I’d like a cuppa”); FORGETTING TO TAKE YOUR PILLS (for a couple of days); and being a bit MUDDLED / CONFUSED (you can see this when someone points it out, but it is UNUSUAL FOR YOU). With specific questions you will say that you are NOT PEEING AS MUCH AS USUAL and having SMELLY URINE although you don’t look at the colour. You have no chest pain and never have. At an appropriate time in the scenario you MUST GO TO THE TOILET, the availability of the urinary dipstick depends on you peeing.

Background: Date of birth – 16 JULY 1933, 76 years old in 2010. 5 FOOT 6 INCHES tall (1.67m) and usually 11½ STONE (73kg). Today 72kg. Usual BMI 26. You have NO KNOWN ALLERGIES. You’re generally FIT & WELL, living ALONE independently, with your cat. Your neighbour comes in most days. Currently you get 2 hours/week of formal home help for domestic chores (‘It’s just lovely having someone come in and wash the floors for me”). You haven’t been in hospital recently. You have NOT had the seasonal influenza jab.
Usual meds: You usually take BETALOC and CARTIA. ONE OF EACH, ONCE A DAY, BUT YOU DON’T KNOW THE DOSES. Very OCCASIONALLY you take a couple of PANADOL but haven’t had any today.

Today’s story: Your neighbour has been away for a few days and on return popped in to see you. Your neighbour found you to be a bit confused. That is, you hadn’t fed the cat, hadn’t collected your paper and you greeted your neighbour with the incorrect name. When your neighbour offered you a cup of tea, you eagerly agreed and drank two cups. Your neighbour was concerned and has brought you to the emergency department at the hospital but has had to leave after talking to the triage
nurse. Right now, as the House Surgeon comes to see you: Your feeling a little thirsty, a bit muddled (“Is this the hospital?”), tired and after walking in from the triage area a little out of breath.
Finding your way around the Otago Virtual Hospital

1. Moving: To move your avatar, use the arrow keys on your keyboard.

2. Seeing: To change your view points, hold the Alt key & click where you want to look (move your mouse to zoom in & out).

3. Chatting: To talk to your team members, type your message into chat bar. Click on Phone (at nurse’s station) to page the registrar or consultant.

4. Examining: To examine your patient, click on patient’s chest, lungs, abdomen, or wrist (note: instruments such as IV line & ECG work in the same way).

5. Prescribing: To prescribe medicines & order lab tests, go to the nurse’s station & click on one of the forms.
Appendix 4: Patient ECG