UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Electronics and Computer Science

Dynamic Feedback Generation in Virtual Patients using Semantic Web Technologies

by

Jean-Rémy Duboc

Thesis for the degree of Doctor of Philosophy

July 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
Electronics and Computer Science

Doctor of Philosophy

DYNAMIC FEEDBACK GENERATION IN VIRTUAL PATIENTS USING SEMANTIC WEB TECHNOLOGIES

by Jean-Rémy Duboc

Virtual patients are interactive tools commonly used by medical schools for teaching and learning, and as training tools for the development of clinical reasoning. The feedback delivered to students is a crucial feature in virtual patients. Personalised feedback, in particular, helps students to reflect on their mistakes and to organise their knowledge in order to use it appropriately in a clinical context. However, authoring personalised feedback in virtual patient systems can become a difficult task, due to the large number of choices available to students and the complex implications of each choice. Additionally, the current technologies used for the design and exchange of virtual patients have limitations in terms of interoperability and data reusability.
Semantic web technologies are designed to model complex knowledge in a flexible manner, allowing easy data sharing from multiple sources and automatic data processing. This thesis demonstrates the benefits of Semantic Web technologies for the design of virtual patients, in particular for the automatic generation of personalised feedback.

Seven important types of personalised feedback were identified from the literature, and a preliminary survey showed that students in year 3 to 5 consider two of these types of feedback to be particularly useful: feedback indicating actions that each student should have chosen but neglected, and feedback indicating the diagnoses that each student should have tested and rule out or confirmed, given the initial presentation of the patient. SemVP, a Semantic Web-based virtual patient system, was created and evaluated by medical students, using a quantitative survey and qualitative interviews. This study showed that SemVP can generate useful personalised feedback, without the need for a virtual case author to write feedback manually, using a semantic model representing both the virtual patient and each student's actions, and leveraging existing data sources available online.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jean-Rémy Duboc , declare that the thesis entitled *Dynamic Feedback Generation in Virtual Patients using Semantic Web Technologies* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:...

Date:...

Acknowledgements

I would like to thank my supervisors for their essential support throughout the three years I spent working on this thesis. I am also grateful to my family, especially Victoria, my wife, who faithfully supported me through the ups and downs of this process. I am also indebted to Dr Nabil Zary and Dr Thomas Nyström, who made my work in Stockholm not only possible but immensely enjoyable.

Finally, I owe a debt of gratitude to all the other people who made this work possible: the clinicians and students who participated in the studies, Scott who helped me with the statistical analysis of the results, and all the students and staff at the university who helped me with the little things of everyday life.

To my niece Valentine Ferey (1995-2011), and to the outstanding medical professionals who fought with everything they had to save her life.

Chapter 1

Introduction

Digital technologies have many benefits for education. For instance, computer systems called virtual patients are used to improve students' ability to apply basic sciences in a clinical setting (Choi et al., 2010). They are also used in order to help students practice their clinical reasoning skills in a safe environment (Poulton and Balasubramaniam, 2011).
This thesis demonstrates how the semantic web, a set of technologies designed to represent complex knowledge across the web, can be used to improve the current state of the art in virtual patients for clinical reasoning. It will be showed in this thesis that a semantic model of virtual patients can generate useful and personalised feedback to medical students automatically, based on the choices they make in a virtual patient scenario.

1.1 Virtual Patient: Definition and Purpose

Ellaway et al. (2006) define virtual patients as "interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment". Users may be learners, teachers, or examiners. This is a broad and inclusive definition, covering a large range of designs and uses, as pointed out by Huwendiek et al. (2009). To specify the pedagogical benefits of virtual patient, two main educational purposes can be identified for virtual patients: situated teaching and learning, and practice of clinical reasoning. The following paragraphs explains how virtual patients can be used for each purpose.

Today's digital tools enable students to access educational information in multiple forms, from multiple locations. Moreover, digital technologies allow interactions with other people (such as fellow students or teachers) and interactive exploration of information. Thus, e-learning is an effective teaching and learning strategy when the tools are designed appropriately and well integrated to the curriculum (McKendree, 2011; Hege et al., 2007; Cook and Dupras, 2004). Virtual patient systems can be a core part of this strategy. For instance, virtual patients can be used for the acquisition of basic biomedical knowledge. In this context, virtual patients are designed as patient-centered teaching tools, guiding the students through a patient's experience, and linking this experience to the underlying clinical processes involved (Choi et al., 2010). This enables situated learning for pre-clinical students. Virtual patients are also used for problem-based learning, a popular method in medical schools (Poulton et al., 2009).

However, theoretical knowledge is not sufficient to become a doctor. Mandin et al. (1997) and Norman (2005) argue that good clinical reasoning doesn't emerge only from generic reasoning skills or superior medical knowledge, but originates from cognitive structures called **schemes** (sometimes also called scripts), elaborated from prototypical or actual patients. Schemes are memory structures used by experts to access relevant knowledge, and to use this knowledge appropriately in context. Schmidt et al. (1990) and Coderre et al. (2003) confirm the importance of scheme-based reasoning in clinical expertise. As a consequence, exposure to many different clinical scenarios is a crucial part of medical education, since it enables the elaboration of clinical schemes. Medical students hone their skills and integrate their pre-existing knowledge by accumulating experience meeting patients with a wide variety of presentations and conditions. These encounters help students to elaborate schemes which they can use in future cases. However, several factors limit the time and frequency of medical students' encounters with real patients, such as the difficulty to obtain patients' consent, budgetary limits, health and safety precautions, rarity of certain conditions, etc.. To enhance students' performance within these constraints, many medical schools use virtual patients as a tool to help students practice their clinical skills (Poulton and Balasubramaniam, 2011; Kenny et al., 2007; Cook and Triola, 2009). Virtual patient systems allow students to practice their diagnostic reasoning skills in a safe environment, while fostering the elaboration of schemes through encounters with simulated patients.

1.2 The Importance of Feedback for Clinical Reasoning Practice

To elaborate useful schemes, students need to reflect on their performance. Larsen et al. (2008) argue that test-based learning centered on feedback about the learner's performance promotes better retention of information. Similarly, Gartmeier et al. (2008) show that negative knowledge (knowledge of what *not* to do) is an important aspect of students' understanding of professional practice. Marshall (1995) also demonstrates that schemes are dynamic structures that should be adapted and modified based on reflective practice.

Thus, feedback is an essential aspect of the learning experience in clinical reasoning.

As a consequence, the delivery of relevant feedback is central to the design of effective virtual patients.

Feedback provides students with information comparing their performance with an agreed standard, in order to improve their abilities (Van de Ridder et al., 2008). To be efficient, feedback must therefore be based on clearly observable skills, must present explicit performance goals, and must be specific about the skills that need improving. Additionally, since feedback is delivered to help each students improve their own performance, efficient feedback needs to be personalised for each student, according to their own performances. In the context of clinical reasoning practice, useful feedback will help each student identify the following mistakes (Friedman et al., 1998; Kassirer and Kopelman, 1989):

- *Faulty Hypothesis Triggering* (failure to identify an appropriate differentiation diagnosis),
- *Faulty Context* (failure to identify relevant aspects of the patient's situation),
- *Faulty Information Processing* (failure to correctly interpret the information gathered throughout the clinical process),
- *Faulty verification* (failure to check or rule out one hypotheses appropriately).

Such high-quality feedback, in particular personalised feedback, can be extremely long and difficult to write, given the level of detail required, and the large number of choices that students can make in a clinical case. A new technological approach is need to facilitate the delivery of useful personalised feedback.

1.3 Semantic Web Technologies: Definition and Benefits

Emerging from the fields of knowledge technology and artificial intelligence, the semantic web provides languages and protocols to represent rich data on the web, perform complex queries and draw inferences across data from various online sources (Berners-Lee et al., 2001). The semantic web enables the creation and reuse of many existing biomedical ontologies and other relevant knowledge bases available on the web, using standard languages such as the **Resource Description Framework (RDF)** and the **OWL Web Ontology Language**.

Virtual patient systems are usually created using data technologies such as relational databases and eXtensible Markup Language (XML[1]). Semantic Web technologies have many benefits over relational databases and XML, in terms of interoperability, expressivity, and data reusability. RDF and OWL enable developers to use complex data from

[1] `http://www.w3.org/TR/REC-xml/`

multiple source in order to describe complex knowledge in a computer-understandable format. This is difficult to achieve using relational databases or XML, as will be demonstrated in Chapter 3.

Using RDF and OWL, it is possible to create a data model that describes both the symptoms affecting the virtual patient and each student's action. Using this data, automatic and personalised feedback can be generated for each student.

1.4 Research Questions

> **This work investigates how semantic web technologies can be used to model virtual patients using existing ontologies and knowledge bases, in order to provide automated and individualised feedback to students based on their interactions with the virtual patient system.**

Such feedback offers the benefit of providing personalised information to each student based on their performance, thus facilitating reflective self-assessment.

To achieve this goal, four main research questions will be addressed:

- What are the most useful types of feedback?
- Do medical students have a consistent understanding and consistent requirements regarding feedback?
- How can semantic web technologies and existing semantic web data be used to generate useful feedback for students, according to their decisions in the virtual patient simulation?
- Does feedback generated using the semantic web improve students' understanding of a virtual patient's case, compared to static feedback?

1.5 Research Scope and Setting

The research described in this thesis was mainly centered on virtual patients used as self-assessment tools for students in Year 3. Indeed, Year 3 medical students at the University of Southampton Faculty of Medicine start their clinical training, and need guidance and training to help them improve their history taking and clinical reasoning skills. Understanding the clinical process involved in various medical specialties is a central learning objectives for this year group, in particular a thorough understanding and practice of the history taking process, built on the knowledge acquired in Year 1 and 2. Therefore, this research was focused mainly on feedback regarding the history taking process. The research conducted throughout this thesis was aimed for students in Year 3 in priority, but was also aimed at students in later years (Year 4 and 5).

The goal of this research was to create a model that would provide useful feedback to students from various cultural and educational backgrounds. To evaluate the model in different educational contexts, the research presented in this thesis was conducted with students from the University of Southampton Faculty of Medicine and with students from the Karolinska Institute in Stockholm, Sweden. Students from both schools start their clinical training in year 3, and therefore have a similar need for training tools in clinical reasoning.

1.6 Document Overview

Chapter 2 provides a general definition of virtual patients, and how they are used in universities from a pedagogical perspective. It will be demonstrated that the design of a virtual patient system is usually guided by its pedagogical objectives.

Chapter 3 is an overview of the semantic web, with a description of the main underlying languages: RDF (Resource Description Framework) and SPARQL (SPARQL Protocol and RDF Query Language) for data encoding and manipulation, RDFS (RDF Schema) and OWL (OWL Web Ontology Language) for vocabularies and ontologies modelling. The benefits of Semantic Web technologies over XML and relational databases are demonstrated, using the Medbiquitous XML specification as an example. Chapter 3 also provides an overview of existing ontologies that can be used to model virtual patients. The choice of a large biomedical ontology called OpenGalen for this research is justified.

Chapter 4 shows the importance of feedback in medical education throughout the literature, in particular for diagnostic reasoning. The most important types of feedback were synthesised in a classification describing 7 different types of feedback. Each type of feedback was rated by medical students during a preliminary study described in Chapter 5. During this study, the requirements for a Semantic Web-based virtual patient system were gathered using a survey and qualitative interviews. During these interviews, a proposed interface model for such a system was evaluated in terms of usability.

Chapter 6 describes the design of **SemVP**, a semantic web-based virtual patient system, based on the results from the preliminary study. The chapter starts with a description of SemVP's interfac design and technical architecture. Then the semantic representations of the virtual patient and of each student's choices are detailed. The final section describes how dynamic feedback is generated from SemVP's underlying model.

A mixed methods study was conducted to evaluate the dynamic feedback delivered by SemVP in comparison to statically authored feedback. Chapter 7 describes the aims of this study, the quantitative experiment design (questionnaires, randomisation), and the qualitative method of enquiry (interview protocols). Chapter 8 contains the results of the study.

Chapter 9 summarises the findings of this research, highlighting the benefits of semantic web technologies for virtual patients demonstrated by the study results, as well as the practical benefits for virtual patient implementation in real learning situations Chapter 9 also presents promising new directions for future research in the field of semantic web technologies for the design of virtual patients.

Chapter 2

Virtual Patients

This chapter presents an overview of virtual patients from the literature, which puts this research in its proper context. The existing interactions and educational approaches afforded by virtual patients systems are explored in more detail. The technologies used to design these interactions are also examined to paint a comprehensive picture of the field.

To start with, it is important to clearly distinguish between virtual patients, virtual cases, and virtual patient systems. It is also crucial to understand the respective roles of the virtual case authors and the virtual case users (in this review, users are medical students).

A virtual patient is a fictional character affected by one or more medical conditions. The virtual patient is represented by digital artefacts such as text (dialogs, descriptions, etc.), still pictures, video, and audio files. These artifacts can represent doctor-patient conversations, lab results, X-Rays, examinations, and any other element that a clinician will use in a clinical context to diagnose and manage the patient's condition (see figure 2.1 for an example).

A virtual case is a clinical scenario presented to students (the virtual patient user) on a computer. A virtual case involves one or more virtual patients, although most cases only involve one patient.

A virtual patient system is the software infrastructure that supports the authoring and delivery of virtual cases. Most virtual patient systems include two components: a virtual case editor (used by virtual case authors) and a virtual case player (used by students).

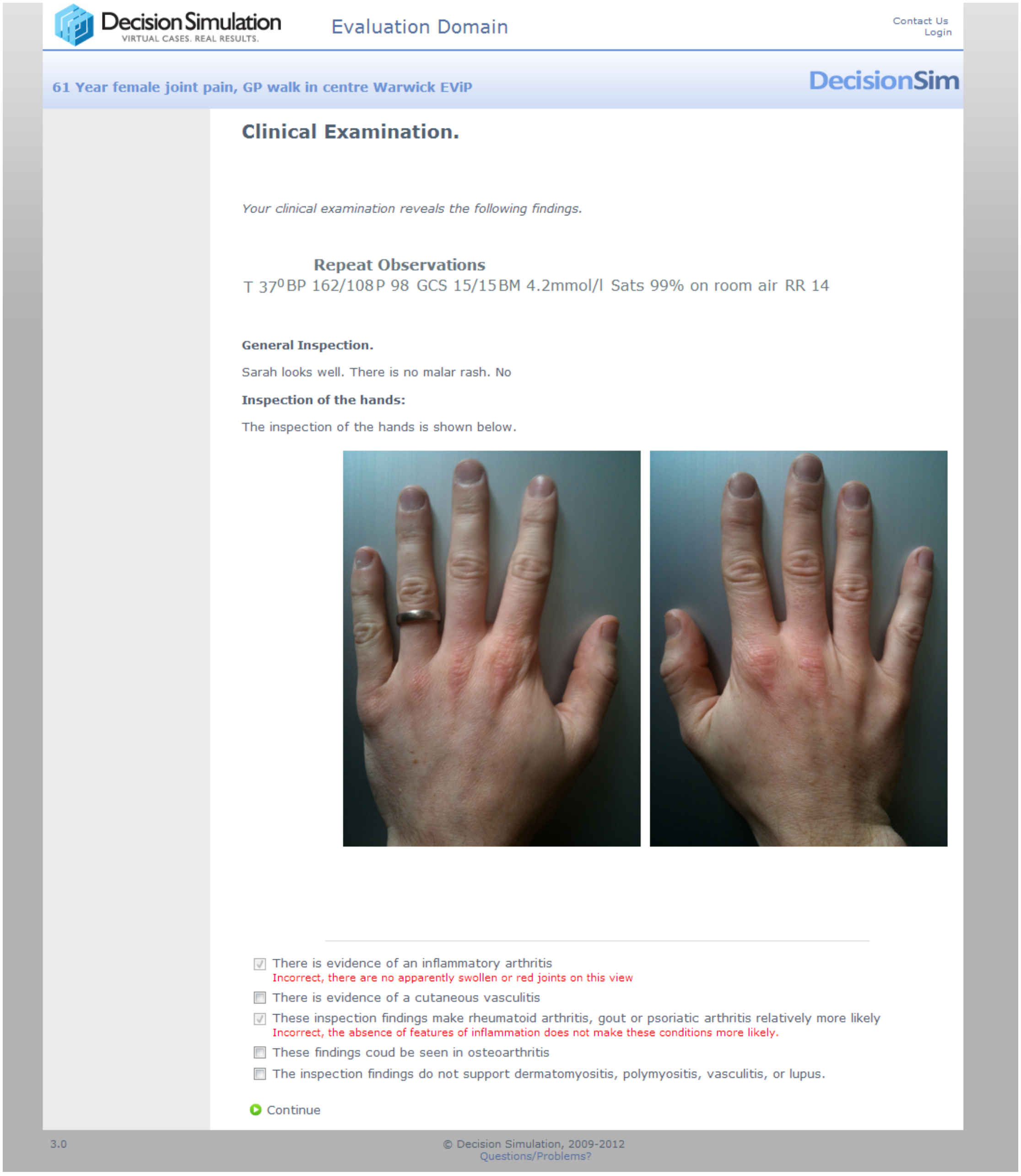


Figure 2.1: Pictures used to illustrate a hand examination in simVP (source: eViP project)

A virtual case author is generally a teacher, who is often also a clinician. The author uses a virtual case editor to design a virtual case, usually by organising all necessary multimedia files. These components can include text, still pictures, audio files, videos, or 3D graphics. Interactive activities (such as multiple-choice questions, clickable pictures, or drag-and-drop exercises) can also be used in some systems. Additional information about the virtual case (such as scientific information explaining the symptoms affecting the patient, references to articles and textbooks, etc.) can sometimes be added at certain stages of the virtual case to support students in the learning process.

Virtual patient authoring represents a significant time investment. The most sophisticated virtual cases have to be designed by multidisciplinary teams comprising domain experts and instructional designers, who oversee the clinical and pedagogical aspects of the case, and multimedia developers responsible for the creation and integration of all multimedia components (video editing, illustration, photos, etc.). These teams require staff with skills in design, computer science, and media production. They bring the storytelling and educational know-how necessary for an engaging and effective learning experience.

The virtual patient user goes through the virtual case and interacts with the virtual patient using the virtual case player. All possible interactions are defined by the virtual case author(s), within the limits of the virtual patient system's features.

Huwendiek et al. (2009) proposed a typology as a common reference language for the study and design of virtual patients. The typology is based on the various virtual cases and virtual patient systems grouped under the project eViP project (electronic virtual patients)[1]. The resulting framework is broad enough to encompass various approaches to virtual patients. It is also precise enough to allow a detailed overview of virtual patients today, and to situate the approach of this research in its broader context.
The typology will be used as a starting point for the overview of virtual patients presented in this chapter, detailing three main aspects of virtual patients:

- Educational: pedagogy is the starting point of any successful virtual patient project.
- Instructional design: once the pedagogic objectives are defined, appropriate instructional design warrants an efficient learning experience for students.
- Technology: virtual patient systems are built using various technologies that support the instructional design and underlying pedagogic needs. A good understanding of these technologies allows a clear view of the potential for future developments in the field of virtual patients systems.

Two main types of educational uses for virtual patients were identified using this framework: virtual patients for basic science teaching and learning, and virtual patients for the development of clinical reasoning. This review shows that virtual patient systems are designed using technologies such as 3D worlds, desktop multimedia software packages, and web technologies. This research is focused on web-based virtual patient systems used for the development of clinical reasoning.

[1] `http://www.virtualpatients.eu/`

2.1 Educational Uses of Virtual Patients and Instructional Design

To insure the success of a virtual patient project, clear pedagogical objectives have to be defined from the onset. The typology proposed by Huwendiek et al. (2009) highlight several key aspects to take into account when designing a virtual patient:

- **Target audience:** who is going to benefit from the virtual patient system? What are their educational needs now?
- **Learning outcomes:** which topic area(s) are covered in the virtual case? What knowledge or skills should the students demonstrate after using the virtual case? How will this be monitored?
- **Learning mode:** is the virtual patient used as a teaching tool, an assessment tool (summative or formative), or a combination of both?

The reviewed virtual patients have been analyzed using this framework, and grouped intro two broad categories, based on the pedagogic objectives of virtual patients.
These categories are:

- Basic science teaching and learning,
- Diagnosis reasoning development.

2.1.1 Virtual Patients for Basic Science Teaching and Learning

Many virtual patients have been used as teaching tools for medicine undergraduates, to teach the core conceptual knowledge underpinning the practice of medicine. In this context, virtual patient are designed as interactive teaching tools meant to replace or complement conventional learning methods such as textbooks, lectures, and paper-based cases. Virtual patients are thus useful tools in the transition from didactic learning to problem-based learning (PBL), an approach pioneered by McMaster University (Saarinen-Rahiika and Binkley, 1998), and increasingly adopted in medical schools. PBL is a teaching method based on exploring and solving problems (usually in small groups) rather than simply acquiring knowledge delivered through lectures and presentations. PBL requires that the problems studied should be too difficult for students to understand with their initial level of knowledge (Schmidt, 1983). The questions that arise about the phenomena described in the problem can be used as learning objectives for students, fostering subsequent self-directed learning. Vernon and Blake (1993) showed that most students prefer PBL over didactic methods of teaching. However, the actual educational benefits of PBL are more difficult to evaluate.
Norman and Schmidt (1992) examined the four assumptions below:

1. PBL fosters clinical reasoning and problem-solving skills,
2. PBL enhances knowledge acquisition, retention and use,
3. PBL improves students' self-directed learning abilities,
4. PBL improves students interest in the subject matter.

After a review of the literature, Norman and Schmidt (1992) found no evidence that PBL improves student's ability to solve problems in a clinical context. However, evidence was found suggesting that PBL may improve knowledge retention on the long term, even though the initial level of knowledge acquired is generally lower in PBL. PBL may also enhance knowledge integration to new problems under certain conditions (Needham and Begg, 1991). Finally, Norman and Schmidt (1992) concluded that PBL does increase students' intrinsic interest in the subject taught and also appears to enhance students' ability to learn independently. Thus, it can be posited that PBL does generally improve students' learning experience in medicine, despite its limitations. This explains why PBL is the pedagogical approach underpinning many virtual patient projects (Poulton et al., 2009; Ruderich et al., 2004; Benedict, 2010). Indeed, virtual patient systems are practical and engaging tools promoting a self-guided or small group exploration of a medical problem, which fits the principles of PBL very well. Poulton et al. (2009) interviewed students at Saint George's University of London (SGUL) after they used paper-based cases and interactive virtual cases. Most students preferred interactive virtual cases to paper-based cases. Interactive features in virtual patient systems promote contextualised learning through storytelling, quizzes and multimedia content such as videos and animations. Storytelling and feedback enable students to reflect on specific medical knowledge and how this knowledge can be applied to solve clinical problems.

Some virtual patients designed for teaching and learning follow a linear or quasi-linear structure, designed to lead students through a clinical case, and allowing for reflection and situated learning through interactive activities. Such virtual patient systems include CASUS (Fischer, 2000), CAMPUS (Ruderich et al., 2004), and virtual cases designed by Choi et al. (2010). Virtual cases presented by Choi et al. (2010), used for year 1 teaching, lead students through the patient's journey from a motorcycle accident to Emergency Department triage, management and long-term recovery (see figure 2.2). Interactive activities are used throughout the case to engage students and encourage them to reflect on key aspects of the patient's journey (figure 2.3). Throughout this scenario, key aspects of the body systems involved in the case are explained through interactive tasks and multimedia learning materials (nervous and locomotor systems). Figure 2.4, for instance, shows activities and video materials embedded in the virtual cases, used in the case to help students learn the Glasgow Coma Scale. Figure 2.5 shows video materials used to demonstrate how to perform a neurology examination. Although not strictly designed for a problem-based learning scenario, this case is a good example of how virtual cases can be used to teach biomedical science, while showing students how these concepts are applied in a clinical context.

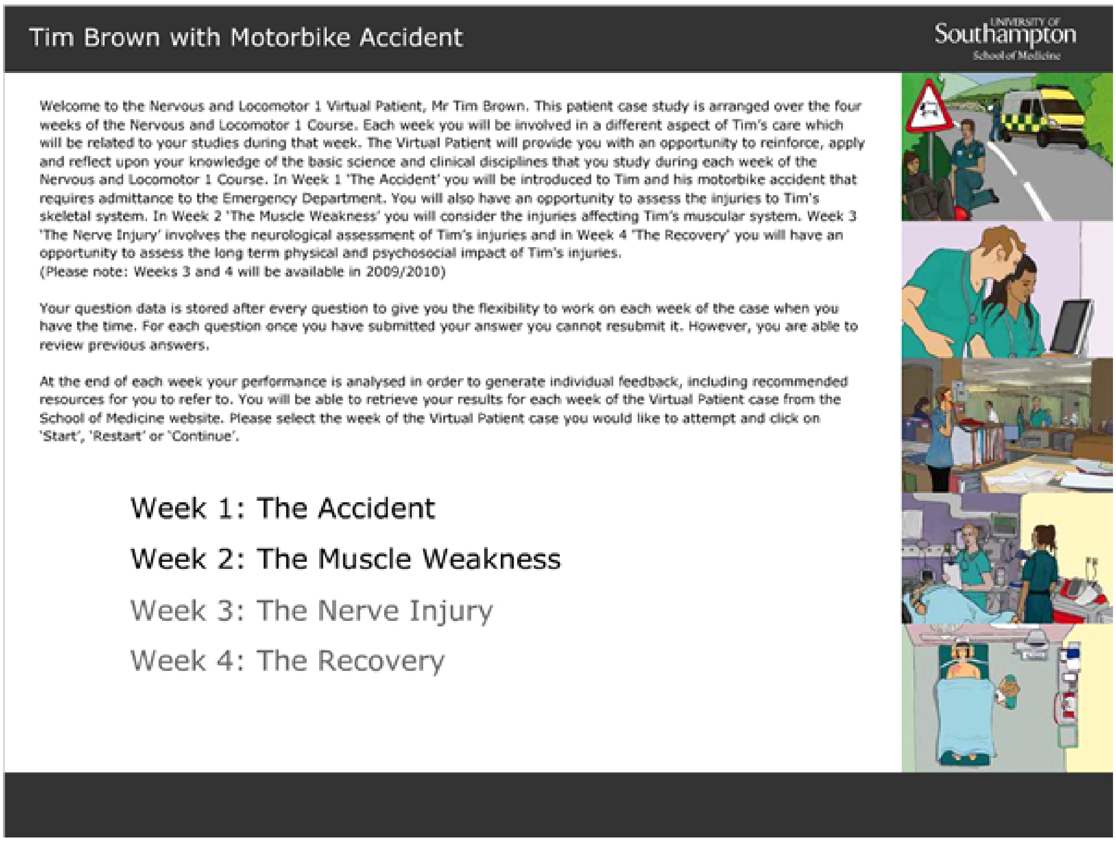


Figure 2.2: Virtual case from Choi et al. (2010): the virtual patient's journey is laid out over a 4-weeks course on Locomotor and Nervous system.

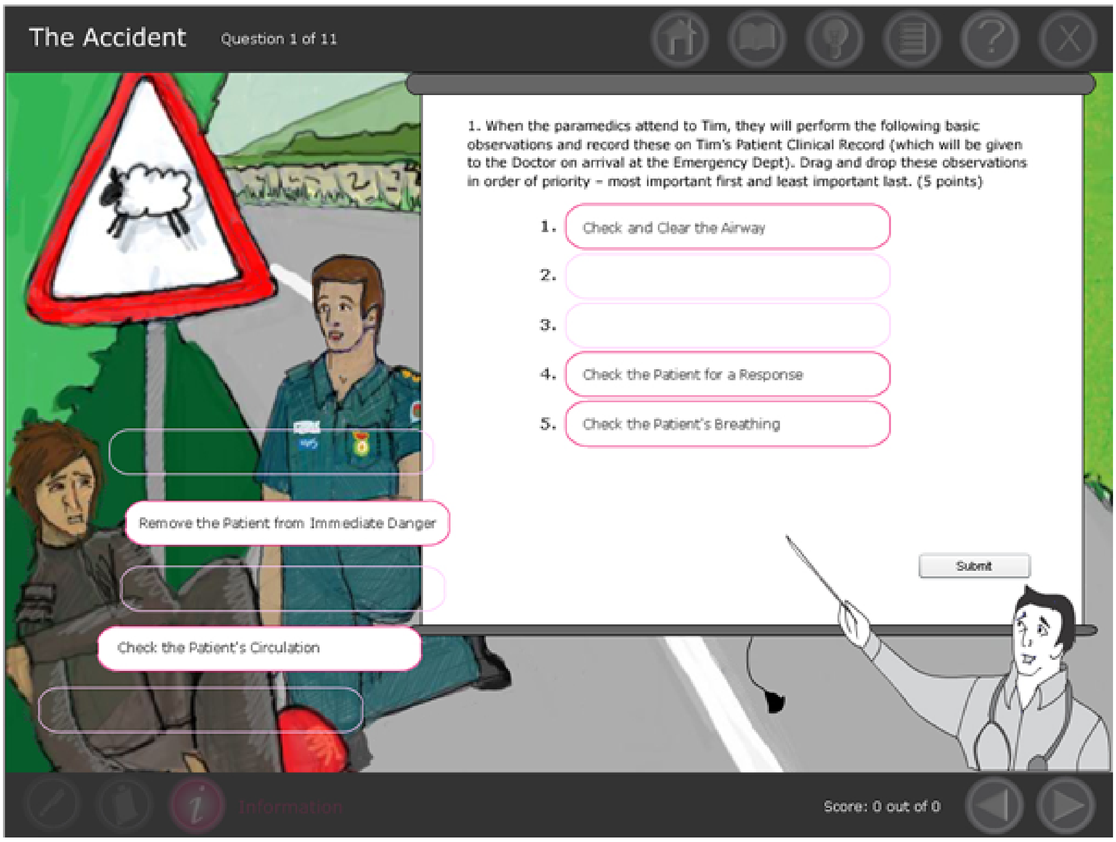


Figure 2.3: Virtual case from Choi et al. (2010): "drag-and-drop" matching exercise designed to help students reflect on the proper course of action to take after an accident.

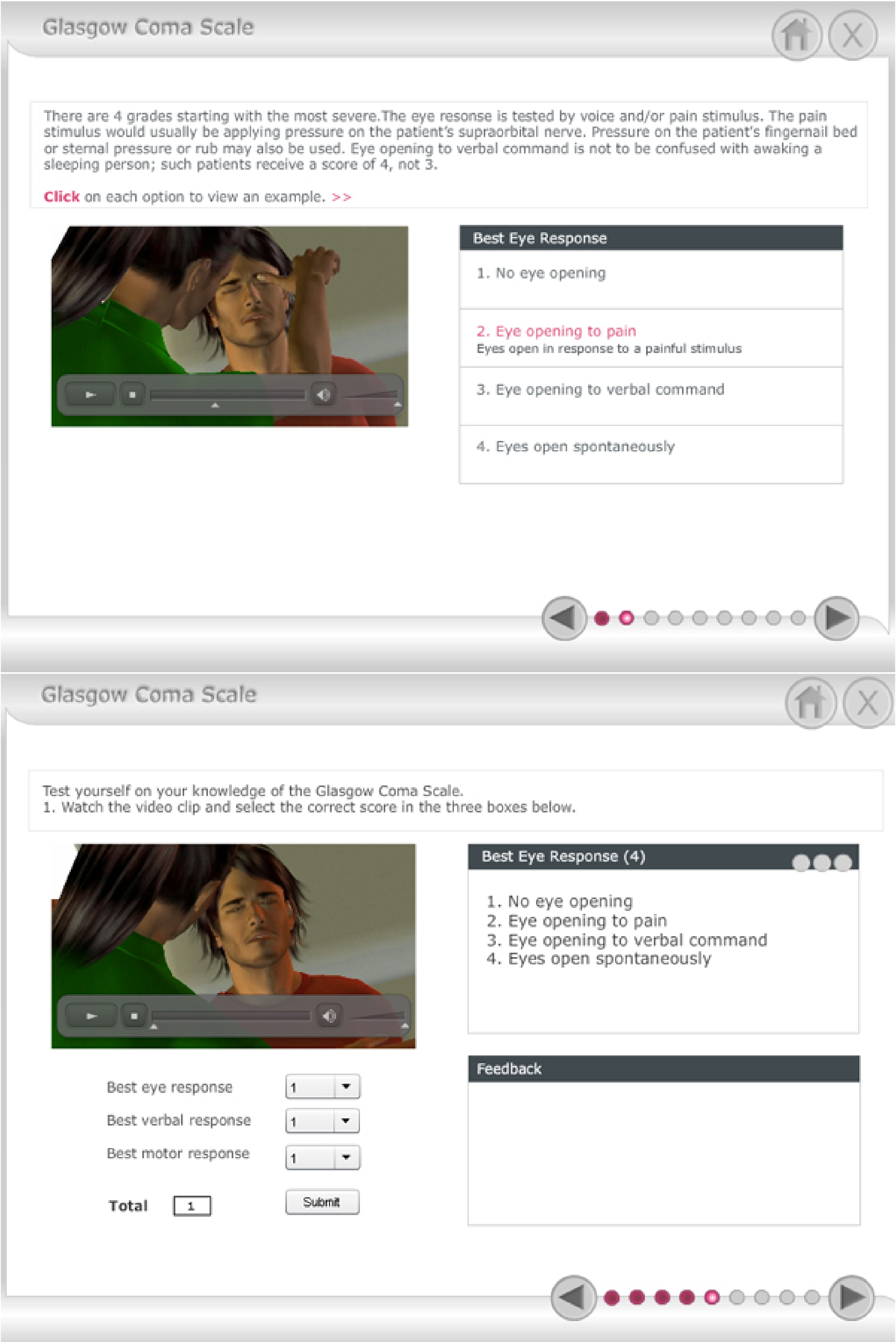


Figure 2.4: Virtual case from Choi et al. (2010): Glasgow Coma Scale (GCS) video resources and activities.

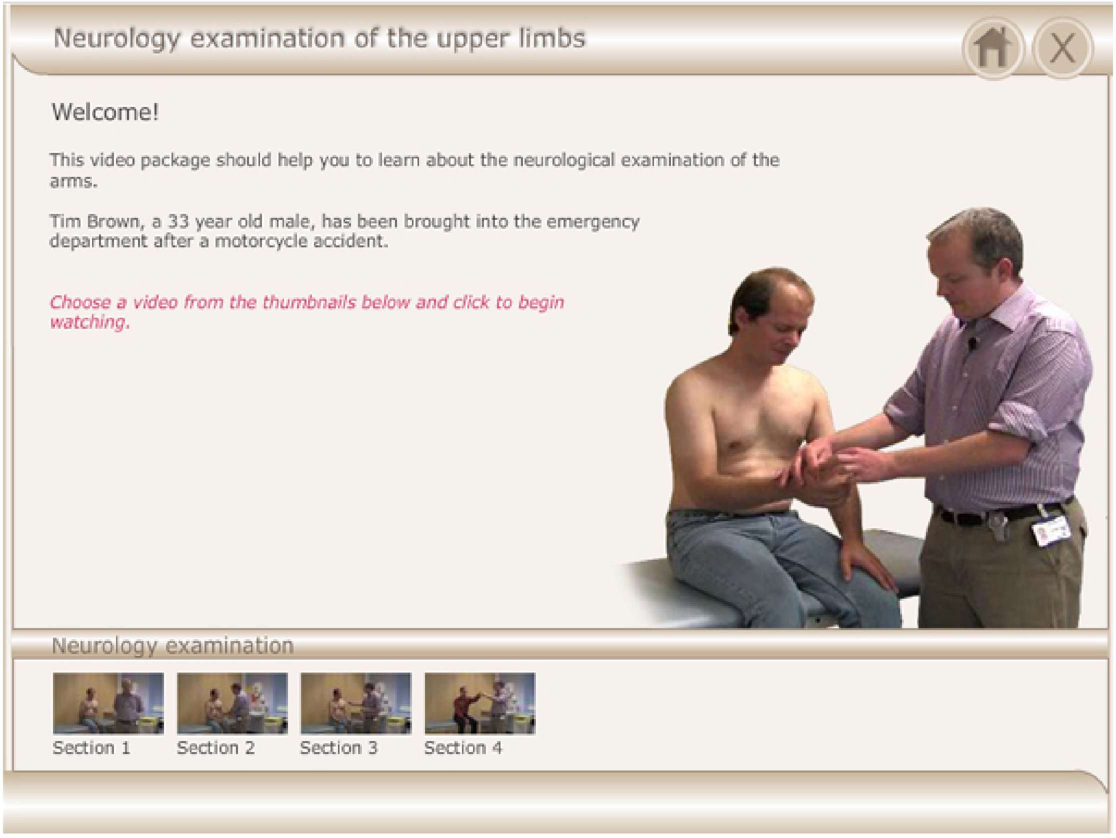


Figure 2.5: Virtual case from Choi et al. (2010): Neurology examination video resources.

Another approach involves using branching paths, and allowing students to explore the consequences of their decisions. OpenLabyrinth (Ellaway, 2010), DecisionSim (Benedict, 2010), vpSim[2], and Quandary[3] are web-based systems designed specifically to create such branching structures. Figure 2.6 shows a branching structure representing a case designed in OpenLabyrinth, and figure 2.7 shows the first page (or "node") of a branching case, as well as a page providing four options for the student to choose from, each leading to a different outcome. Labyrinth, Quandary and vpSim have been compared in terms of usability by Sawdon and Curtis (2010). No conclusion was reached about the educational impact of the tools in that study, however.

Poulton et al. (2009) describes the use of OpenLabyrinth as a replacement and enhancement of traditional paper-based patient cases. The conclusion was that branching virtual patients are generally more engaging than linear cases. Branching cases allow students to explore various options and reflect on the consequences. Additionally, expert feedback is provided when appropriate, depending on each student's choice while exploring the case (figures 2.8 and 2.9 show feedback delivery in two virtual patient systems). This feedback is an essential part of the learning process. Indeed, an experiment conducted by

[2]`http://vpsim.pitt.edu/shell/Login.aspx`
[3]`http://www.halfbakedsoftware.com/quandary.php`

Needham and Begg (1991) showed that failure to provide feedback in a problem-based learning scenario can reduce or even eliminate students' ability to apply their knowledge to future problems.

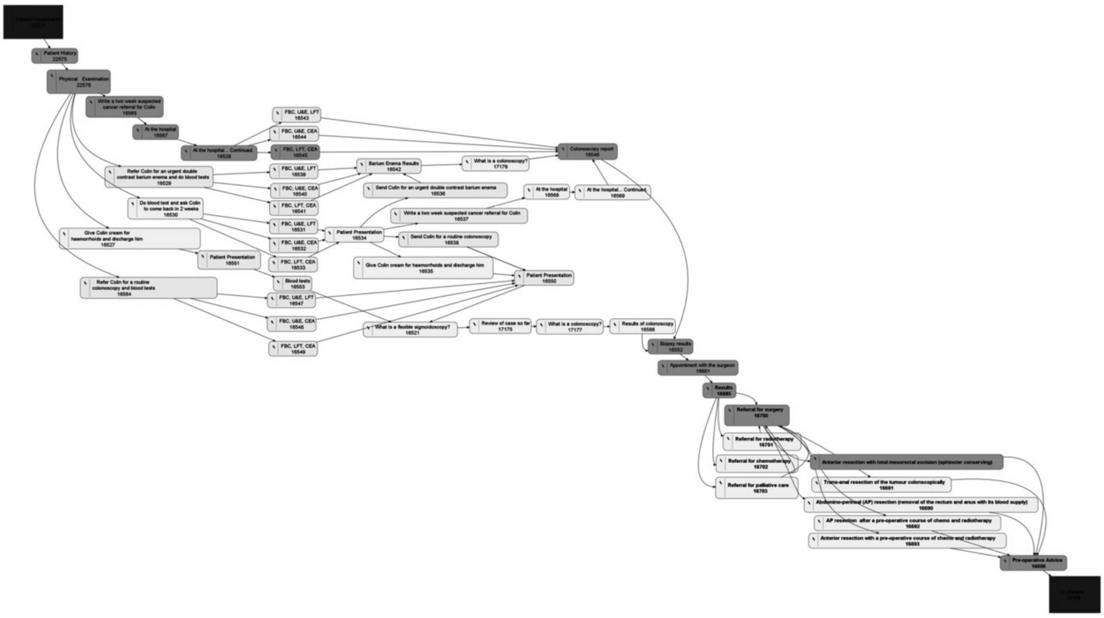


Figure 2.6: Virtual case branching structure (source: Poulton et al. (2009))

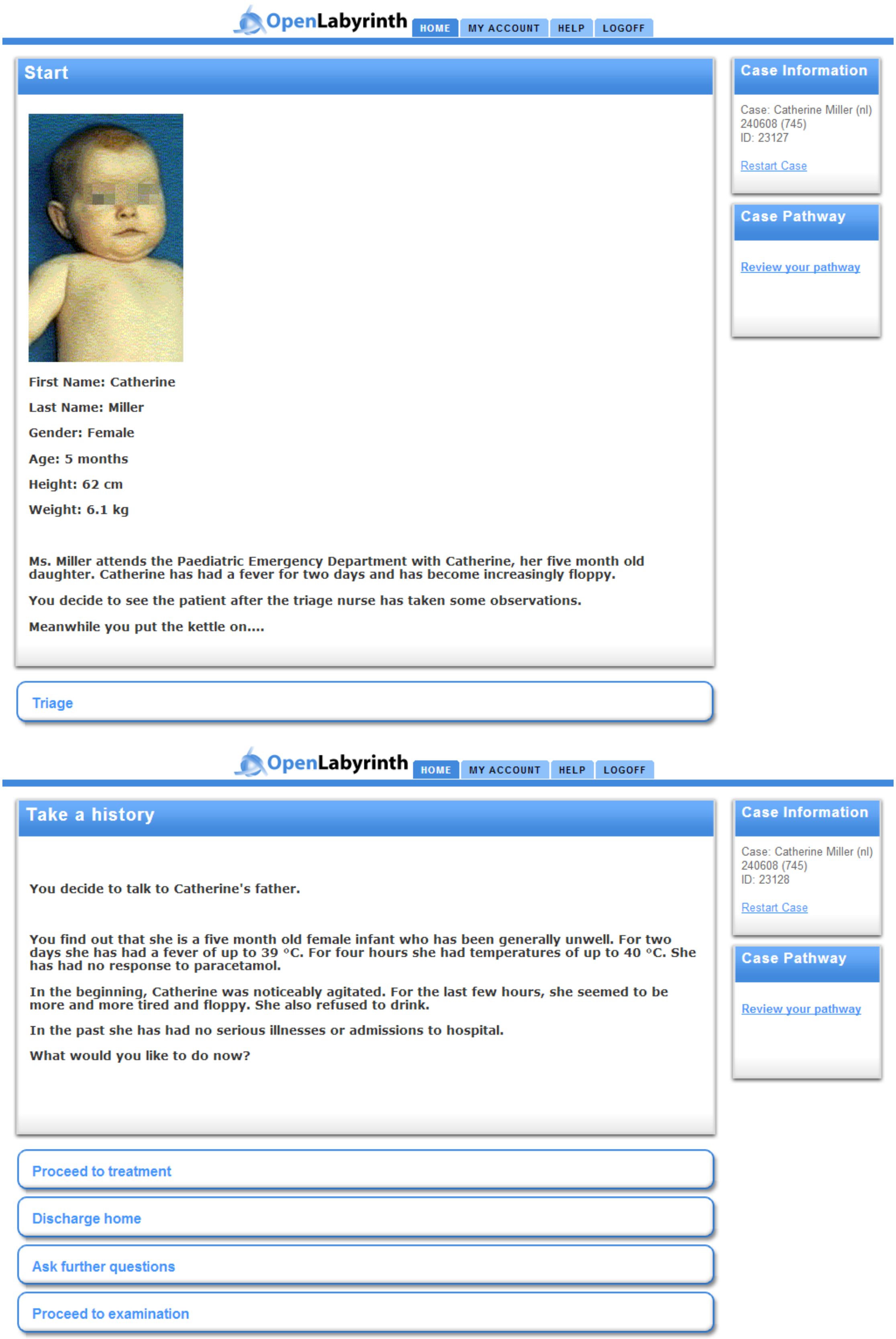


Figure 2.7: Screen captures from a branching case designed in OpenLabyrinth (source: eViP project)

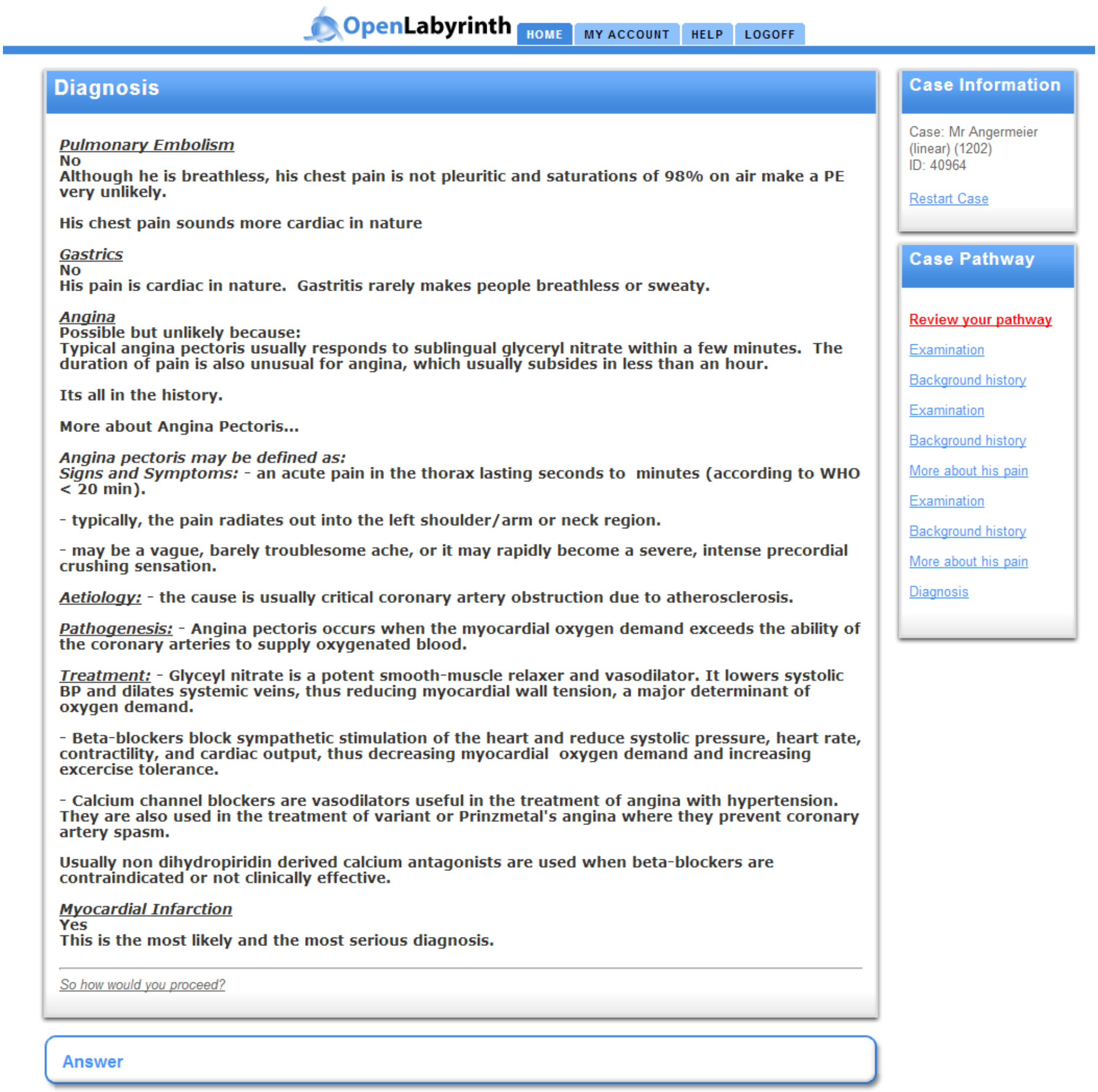


Figure 2.8: Feedback provided on diagnosis choices in OpenLabyrinth (source: eViP project)

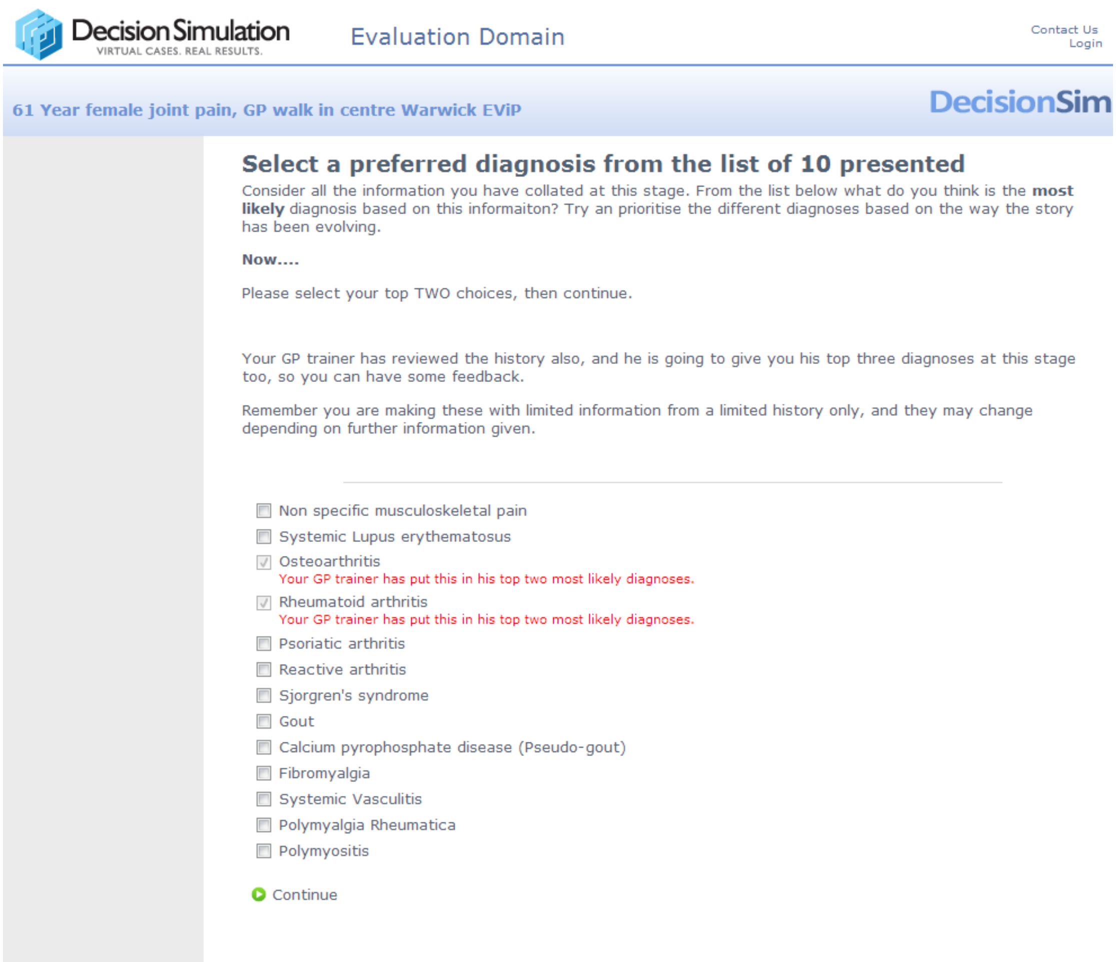


Figure 2.9: Feedback provided on diagnosis choices in vpSim (source: eViP project)

Branching cases are useful for problem-based learning, and they can also be beneficial for the development of clinical reasoning.

2.1.2 Virtual Patients for The Development of Clinical Reasoning

For students in later year groups starting clinical training, biomedical science knowledge alone is not sufficient; future doctors need to apply their knowledge in a clinical context to diagnose and manage patients appropriately. Virtual patients can be used to help students achieve this goal by improving their clinical reasoning through problem-based learning. Many existing virtual cases were designed to support the development of clinical reasoning (Fischer, 2000; Gozum, 1994; Lyon et al., 1992). Cook and Triola (2009) also asserts that virtual patients are well suited for this purpose. This research shares this focus on the improvement of clinical reasoning through virtual patients.

Virtual patient systems allow students to explore many different cases, which helps them understand how medical knowledge should be applied in a number of realistic situations.

Virtual patient systems also allow students to encounter clinical cases they may not have the opportunity to see during medical placements (rare diseases, unusual presentations, etc.). Finally, virtual patients systems present the tremendous advantage of delivering reusable and standardised cases containing carefully written expert feedback, which promotes consistent practice for all students.

Virtual cases designed for the development of clinical reasoning in mind usually enable students to interact with the patient in a non-linear manner, by selecting or typing questions, choosing examinations and lab tests, proposing their diagnoses, and choosing management options (Toro-Troconis et al., 2008; Fischer, 2000). Branching structures have been used for this purpose. Branching structures can be used to encourage students to explore the consequences of their choices on a patient, which provides valuable feedback. Less restrictive structures called "exploratory" cases have also been used. Figure 2.10 shows the exploratory path provided in Web-SP (Zary et al., 2006). Exploratory cases have the benefit of being realistic, and enabling students to choose actions to perform freely, as they would do in a real clinical situation. SemVP (presented in Chapter 6) is also an exploratory virtual patient system.

To support their learning, students receive feedback based on the choices they make. Feedback enables students to reflect on their actions and to understand their mistakes in a clinical situation. Feedback is indeed a crucial feature for the development of clinical reasoning, as will be demonstrated in Chapter 4.

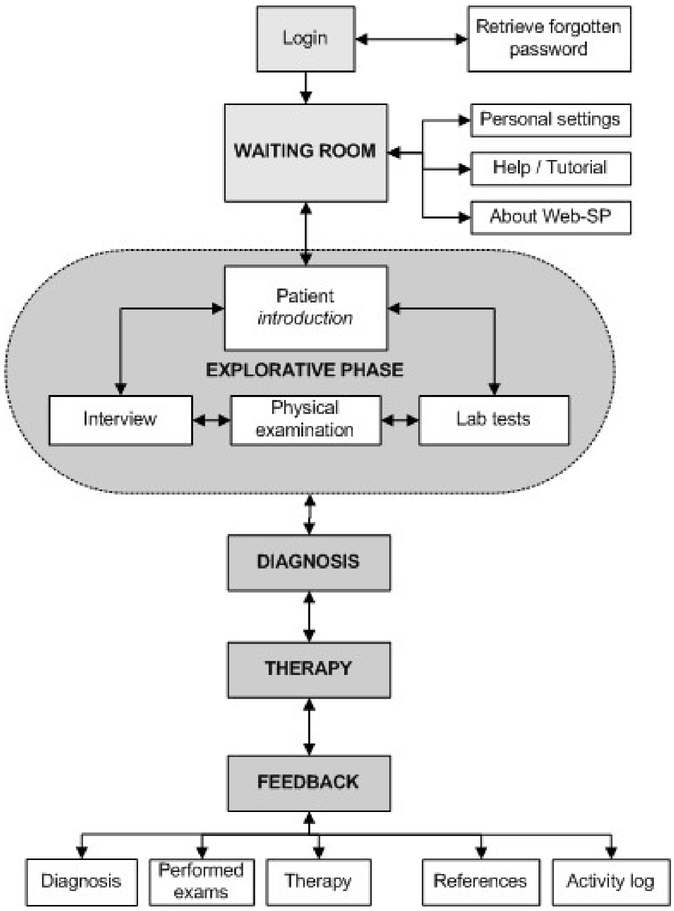


Figure 2.10: Exploratory structure of Web-SP (source: Zary et al. (2006))

2.2 Virtual Patients' Integration to the Curriculum

The success of using virtual patients to facilitate learning depends heavily on a good integration to the medical curriculum. Proper curriculum integration fosters good adoption rates by students and teachers, better sustainability, and improved learning experiences.

The work conducted by Hege et al. (2007) presents evidence that case-based e-learning application such as virtual patients tend not to be used by students unless they are integrated to the curriculum as compulsory assessments, or contribute directly to the learning objectives of the student. Using virtual patients to prepare for examinations such as the Objective Structured Clinical Examination (OSCE) and the Mini Clinical Evaluation Exercise (Mini-CEX) is a strong motivator (Sloan et al., 1996; Casey et al., 2009).

Botezatu et al. (2010) confirms the role of virtual patients as personal revision tools for clinical reasoning through qualitative data gathered from a small group of 8 students on the use of Web-SP (Zary et al., 2006). Memorisation, transferable skills and evaluation of mistakes (feedback) are also mentioned as important benefits of virtual patients. Five themes for a successful virtual patient are identified by the authors: learning, teaching, assessment, authenticity and implementation. These results confirm the importance of an adequate integration of virtual patients to the curriculum, as established by Hege et al. (2007). However, this study has been conducted with a limited number of participants, and the results need to be supported by other similar research. Poulton and Balasubramaniam (2011) note that virtual patient are now increasingly used in medical schools, as their educational value becomes clearer for institutions. They are also better integrated to the curriculum, to the point of actually driving change in curricular design.

2.3 Technological Approaches

Before the democratisation of the web, virtual patients have been designed using desktop multimedia technologies (Lyon et al., 1992; Fischer, 2000; Lyon and Fisher, 2001).
The widespread use of web technologies in the last 10 years opened up a wide range of possibilities for virtual patient systems. Many current systems use web technologies because of their numerous benefits (Fischer, 2000; Zary et al., 2006; Begg et al., 2007). These benefits include:

- Users can access the application from anywhere without any software installation or upgrade, on most operating systems,
- Web-based virtual patient systems can scale to thousands of users with a single code-base,
- All upgrades to the system or to virtual cases can be deployed to every user in real time,
- Well-designed web platforms can become efficient collaborative environments, allowing students to communicate about the cases with their peers and teachers,
- Data describing users' activities is recorded on a central server and can be analysed for technical and pedagogic purposes.

However, Zary et al. (2006) also point out obstacles to the adoption of web-based systems in health education, such as the absence of a common standard and generic platform for the creation and management of virtual cases, and the dependence on computer specialists to support cases creation and maintenance. The lack of tools to exchange cases between teachers and systems is also an issue, which the Medbiquitous XML standard is aiming to solve (see section 3.5.2 for more details on Medbiquitous XML and the benefits of the semantic web over this specification).

These technologies offer great potential for designing and delivering virtual patients. However, they present limitations in terms of graphic display capabilities. Indeed, at the moment the web mainly affords text and 2D image display. Using videos and audio content is possible with additional plug-ins such as Adobe Flash Player[4] or natively on some browsers using the recent HTML5[5]. Some systems such as the Imperial College Second Life hospital (Toro-Troconis et al., 2008) or PIVOTE[6] are built using 3D virtual worlds. Virtual worlds may create more immersive learning experience, even including elements designed to trigger an emotional response similar to what may occur in a real clinical situations (Cavazza and Simo, 2003). However, Cook and Triola (2009) and Merriënboer et al. (2002) suggest that the graphic realism of the simulation is less important than the quality of feedback and the development of mental models. Experience shows that the success of virtual patient projects depends more on the underlying design and pedagogic work than on the technology supporting it. Therefore, the current limitations of the web in terms of graphics may not be an obstacle to the creation of effective web-based virtual patients, and learning design may be the most important feature for an engaging and efficient learning experience.

Virtual patients seem to be ideally suited for the development of clinical reasoning, through the deliberate practice of a variety of cases. This can be achieved using various technologies, but web-based systems offer a flexibility that make them ideally suited for this type of application.

[4]`http://www.adobe.com/products/flashplayer/`
[5]`http://www.w3.org/TR/html5/`
[6]`http://code.google.com/p/pivote/`

Chapter 3

The Semantic Web

This chapter presents a broad overview of the Semantic Web, a group of technologies designed to make data published on the Web easily accessible across locations and applications, and thus to move from a web of document to a web of data. A short history of the Semantic Web is presented, along with an overview of the languages underpinning the Semantic Web: RDF and SPARQL for resource description and retrieval, and RDFS and OWL for ontology design.

The benefits of the Semantic Web over other technologies are discussed. Existing specifications and ontologies that present a potential for the design of virtual patients are reviewed. The OpenGalen ontology, associated with ontology design patterns, was chosen as a foundation for a Semantic Web-based virtual patient system.

3.1 The Origins of the Semantic Web: from a Web of Documents to a Web of Data

The first web page was available in 1990, on a system designed by Sir Tim Berners-Lee, in collaboration with Belgian computer scientist Robert Cailliau. Originally, the Web was designed as a document sharing system, created essentially to cope with the different document formats that scientists were using at CERN (Centre Européen pour la Recherche Nucléaire, or European Organization for Nuclear Research[1]). It was built on top of the Internet, a decentralised "network of networks" used by the scientific community to exchange information (Berners-Lee, 1999). The Web is now made up of billions of individual electronic pages, linked together by hyperlinks. However, if human beings can easily read and understand these documents, the sheer number of web pages makes finding information manually extremely difficult. One solution to this problem is the design of web search engines; one of the most famous search engines, Google, was

[1] `http://cern.ch`

created by Sergei Brin and Lawrence Page in 1998 (Brin and Page, 1998), and remains an essential tool for finding information on the web. Indeed, search engines have been an important factor in the development of the Web itself, since they enable users to find information on an exponentially growing network of web pages. However, search engines are susceptible to problems such as low precision of results due to ambiguous vocabulary in web pages. These difficulties arise because web pages contain data in natural language, and not in formats and languages designed to query, filter and combine data using a computer program (Antoniou and Van Harmelen, 2008).

The Semantic Web is a solution designed to share and link information on the Web in a way that makes it easy to query and combine automatically using computer programs. Using reasoners (computer programs capable of performing simple logical operations), it is also possible to make automatic inferences based on pre-existing information, i.e. to deduce new information from pre-existing data/.

The Semantic Web is a set of technologies built on top of each other in "layers" (see figure 3.1), which facilitate the automatic retrieval and processing of information available on the web. It leverages existing protocols such as Hypertext Transfer Protocol (HTTP) (Fielding et al., 1999) and eXtensible Markup Language (XML) (Bray et al., 2008). Its first building block is the Resource Description Framework (RDF) (Manola and Miller, 2004), described in more detail in Section 3.2. SPARQL Protocol And RDF Query Language (SPARQL), described in Section 3.3, is a protocol allowing users and computer programs to query RDF data from a remote server on the web (Clark et al., 2008) and a query language designed to retrieve information written in RDF (Prud'hommeaux and Seaborne, 2008). RDF Schema (RDFS) and the Web Ontology Language (OWL), built on top of RDF, are designed to describe classes of resources and how these classes relate to each other (Brickley and Guha, 2004; McGuinness and van Harmelen, 2004). These languages are used to represent how resources are organised in specific domains, for instance medicine or biology. RDFS and OWL will be described in more detail in Section 3.4.

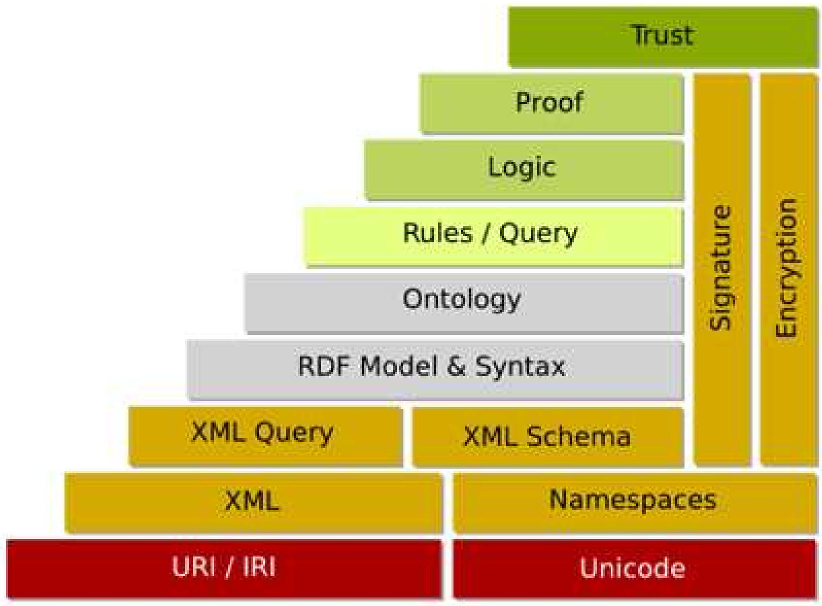


Figure 3.1: The semantic web "layer cake"

3.2 Representing Resources on the Web: the Resource Description Framework (RDF)

RDF is a languages designed to make simple declarations about *resources*. Most commonly, resources are documents and files that can be retrieved on the web, such as web pages and downloadable files. Information about such Web resources can include, for instance, title, author, modification date, and copyright information. The notion of a resource can also be generalised to things that are *represented* on the web, but not necessarily retrieved directly on the web. This includes, for instance, physical products available from an online shop, people belonging to an organisation, and of course virtual patients.

Each resource is identified using a *Universal Resource Identifier*, or URI. It is possible to make statements in RDF about anything that is identified by a unique URI. Each statement is represented as a *triple*, comprising a *Subject*, a *Predicate*, and an *Object*. For instance, in the statement "Virtual Patient X has a creator who is John Y", the subject is "Virtual Patient X", the object is "John Y", and the predicate is "creator". The subject identifies the resource described in the statement (Virtual Patient X), the predicate describes a property of the resource (creator), and the subject defines the value of this property (John Y). In RDF, subjects and objects are represented as nodes, and predicates are represented as arcs, so that each statement is represented as an oriented graph. Any additional statements can be made about *John Y* or about *Virtual Patient X*, resulting in a growing graph of interconnected data. Since every resource has a unique identifier (URI), complex graphs of interconnected statements about numerous resources can be created. For instance, the statement "John Y's workplace home page is the University of Southampton's home page" can be added to the graph, with John Y as the subject, "workplace home page" as the predicate, and University of Southampton's home page URL as the object. These two statements are represented as a graph on figure 3.2.

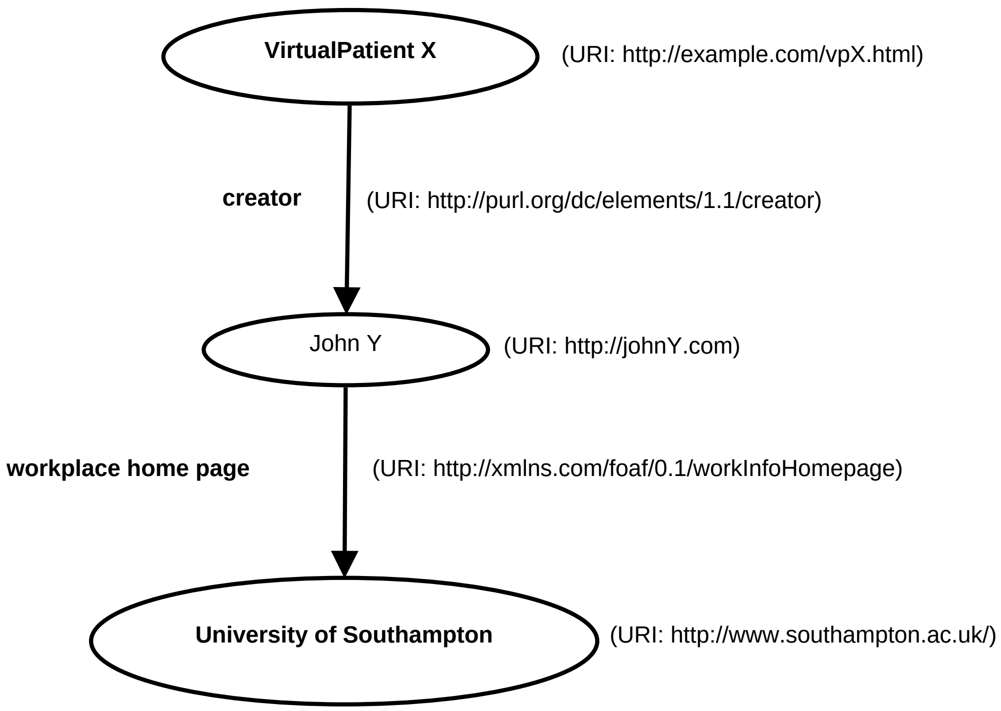


Figure 3.2: Two simple statements in an RDF graph. Full URIs are written next to the corresponding nodes and arcs.

In the case of a virtual patient, many statements about the patient can be represented in RDF, such as "Patient X is a male patient" or "Hand pain is a symptom of arthritis". This is a first step towards a machine-readable model of virtual patients. Using RDF, it is possible to represent virtual patients as graphs of connected data, and to make automatic inferences(deduce new information) based on this data. Chapter 6 demonstrates how this can be achieved.

Literal Nodes

Since objects in statements represent values for a given property, it is possible to give them literal values, such as strings or numerical values. For instance, the string "Mr. John Y"can be used as a value to the property "family name". The corresponding statement would read as follows: "John Y has a family name whose value is "Mr. John Y"". Literals can have a number of types, including the built-in datatypes already available in XML (Biron and Malhotra, 2004). These datatypes include variables indicating time (*duration*, *dateTime*, *date*, etc.), numerical variables (integers, boolean, decimal, etc.) and more. Conventionally, literals are represented in rectangular boxes on graphs (see figure 3.3).

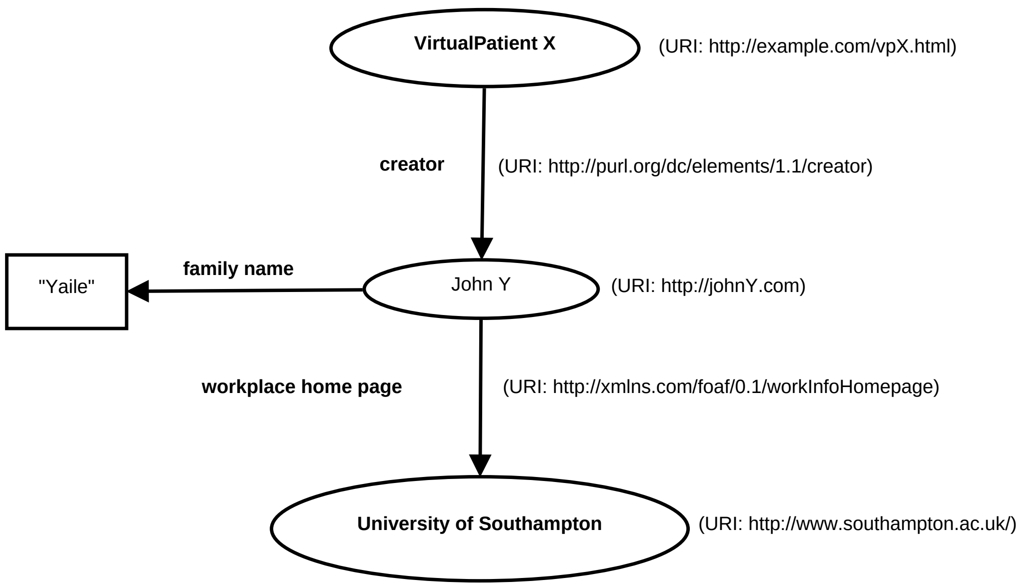


Figure 3.3: Blank literal used to represent John Y.'s family name

Namespaces

To simplify notation further, it is possible to designate a resource without writing the full URI, using a shorthand. The shorthand substitutes an XML qualified name (QName) for a reference to a base URI. The complete URI is then reconstituted using the base URI and the rest of the URI. The full URI for RDF itself is `http://www.w3.org/1999/02/22-rdf-syntax-ns`, and is typically replaced by `rdf:`. Thus, the statement below (in plain English: "Patient X is a Virtual Patient"):

```
<http://example.com/patient_x> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://example.com/virtualpatient>.
```

can be shortenened as follows:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

<http://example.com/patient_x> rdf:type  <http://example.com/virtual_patient>.
```

The whole triple can be shortened even further as follows:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex: <http://example.com/>.

ex:patient_x rdf:type  ex:virtual_patient.
```

3.3 Retrieving and Modifying RDF Data: SPARQL

RDF data is only useful if tools are available to retrieve and modify it. SPARQL (SPARQL Query Protocol and Language for RDF) is a language and protocol designed for this purpose. SPARQL allows computer programs to access RDF data from a remote server (Prud'hommeaux and Seaborne, 2008). As a querying language, SPARQL is designed to match the data queried to a set of *triple patterns* called *basic graph pattern*. Triple patterns are very similar to RDF triples, with the exception that triple patterns can contain variables as subject, predicate, or object.

For instance, a query returning all female patients in a knowledge base could read as follows:

```
SELECT ?patient
WHERE
{
        ?patient rdf:type opengalen:femalePatient.
}
```

This query can be translated in English as follows: "select all nodes which belong to the class *femalePatient*". A slightly more complex query returning all female patients' names would read as follows:

```
SELECT ?patient ?name
WHERE
{
        ?patient  rdf:type                          opengalen:femalePatient.
        ?patient  <http://xmlns.com/foaf/0.1/name> ?name.
}
```

The results would be returned as data bindings, as represented in the following table:

Patient	Name
<http://example.com/vps/jane_foster>	"Jane Foster"
<http://example.com/vps/ms_matibunda>	"Ms. Matibunda"
<http://example.com/vps/catherine_m>	"Catherine M."

SPARQL query results can be filtered and ordered. Section 6.5 shows how result filtering can be used to generated automatic feedback. SPARQL can also be used to add new triples to RDF graphs using the INSERT query form. The following query will add a triple representing a new patient called Michael B.:

```
        INSERT {<http://michaelB.com> rdf:type opengalen:malePatient.}
```

3.4 Representing Domain Vocabularies: RDF Schema and the OWL Web Ontology Language

RDF is a useful tool to make statements about resources, but it lacks the ability to define the terms (vocabulary) used in those statements. This gap is filled by **RDF Schema** (or RDFS) (Brickley and Guha, 2004). RDFS is essentially a set of RDF resources with special meaning. Using the predefined resources available in RDFS, it is possible to organise classes (or types) of resources in a hierarchical structure, a vocabulary. It is also possible to designate some resources as properties and to define how these properties relate to certain classes. RDFS thus allows the creation of simple vocabularies (or schemas), designed to model a given domain of knowledge. Computer programs, called RDFS reasoners, can then use these vocabularies to make simple inferences about resources, deducing new information given existing data. RDFS is a practical tool for the design of a semantic model for virtual patients, in that it affords the creation of a common vocabulary for all virtual cases. **OWL (Web Ontology Language)** extends RDFS and provides means to describe more complex relationships between classes and their properties, thus allowing the design of domain ontologies. Many existing biomedical ontologies are designed in OWL; classes and properties from these ontologies can be used in a semantic model of virtual patients The following sections provide an overview of RDFS, and a short introduction to OWL.

3.4.1 Defining a Class Hierarchy in RDF Schema

An RDF Schema class is simply any RDF resource with `rdfs:Class` as value for its `rdf:type` property. Using an example from the previous section, the resource `ex:VirtualPatient` can be defined as a class, using the following statement:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex: <http://example.com/>.

ex:virtualpatient rdf:type  rdfs:Class.
```

Subsequently, all resources who have `ex:VirtualPatient` as value for their `rdf:Type` property belong to the RDFS class `ex:VirtualPatient`. They are called *instances* of the class. A resource can be an instance of several classes. A resource representing a virtual patient can belong, for instance, to the `ex:VirtualPatient` class as well as the `gender:Male` class.

RDFS also provides a resource called `rdfs:subClassOf`. This resource describes a relation of subsumption between two classes: one class is a sub-group of another. For instance, a class called `ex:Human` could contain two subclasses called `ex:Man` and `ex:Woman`. The triples describing this relationship reads as follows:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex: <http://example.com/>.

ex:Human        rdf:type        rdfs:Class.
ex:Man          rdfs:subClassOf ex:Human
ex:Woman        rdfs:subClassOf ex:Human.
```

`rdfs:subClassOf` is a transitive property, which means that instances of subclasses are also members of classes higher in the hierarchy. In the previous example, an instance of `ex:Man` is *de facto* an instance of `ex:Human`.

3.4.2 Describing Properties in RDFS

Properties in RDF are defined in a similar way to RDFS classes: an RDF Schema property is any RDF resource with the RDF class `rdf:Property` as value for its `rdf:type` property. The `rdfs:range` property is used to indicate that the values of a particular property are instances of a designated class. For instance, the property `ex:hasDaughter` should have only instances of `ex:Girl` as values, because a daughter can only be a girl. Conversely, it is possible to establish that, in the specific domain described, someone can only be the daughter of a human being. It is possible to do so using the `rdfs:domain` property, which specifies that a given property applies to instances of a designated class. This is described as follows:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex: <http://example.com/>.

#define the 2 classes ex:girl and ex:Human
ex:Girl         rdf:type        rdfs:Class.
ex:Human        rdf:type        rdfs:Class.

#define the property ex:hasDaughter
ex:hasDaughter  rdf:type        rdf:Property.

#indicate that the values ex:hasDaughter should be instances of ex:girl
ex:hasDaughter  rdfs:range      ex:Girl.
#indicate that the property ex:hasDaughter should apply to instances of ex:Human
ex:hasDaughter  rdfs:domain     ex:Human.
```

RDFS properties can also be subsumed in the same way classes can be subsumed, using the `rdfs:subPropertyOf` property. Using the previous example, `ex:hasDaughter` can be defined as a sub-property of `ex:hasChild`.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex: <http://example.com/>.

#define the properties ex:hasDaughter and ex:hasChild
ex:hasDaughter  rdf:type        rdf:Property.
```

```
ex:hasChild      rdf:type        rdf:Property.

#define ex:hasDaughter as a sub-property of ex:hasChild
ex:hasDaughter  rdfs:subPropertyOf      ex:hasChild
```

This allows inferences based on the transitivity of `rdfs:subPropertyOf`. In this example, a resource defined as the daughter of another resource (using `ex:hasDaughter`) will also implicitly be defined as the child of the same resource (using `ex:hasChild`).
Figure 3.4 illustrates the vocabulary described in this section.

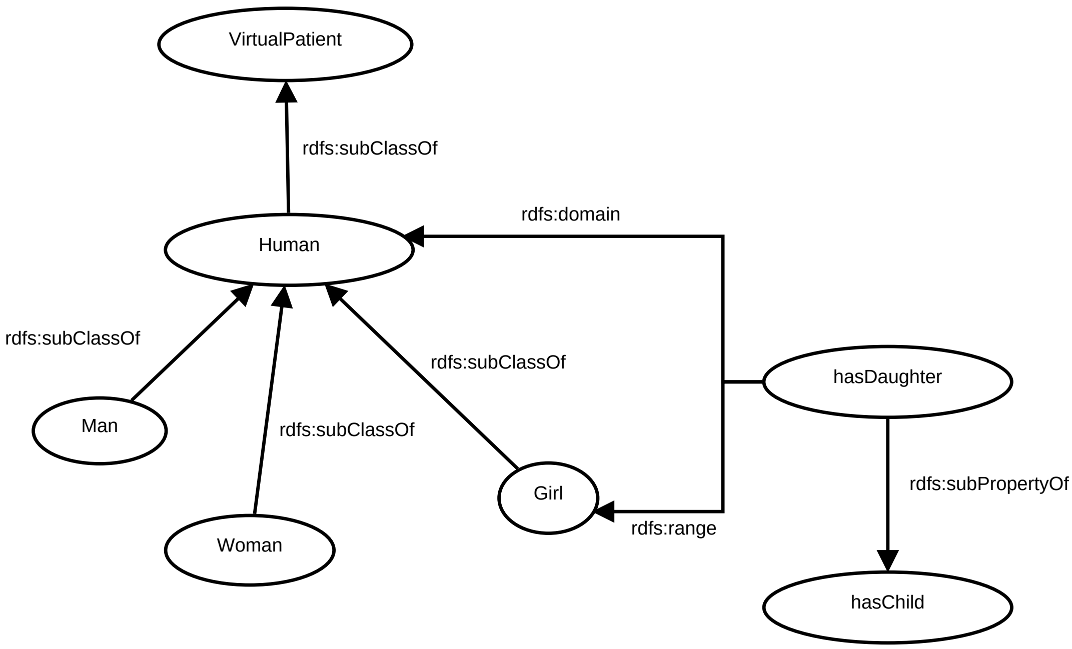


Figure 3.4: A example vocabulary in RDFS

RDFS provides a few other resources with special meaning, such as:

- `rdfs:Resource`, the top-level class for all resources,
- `rdfs:Literal` and `rdfs:XMLLiteral` for data types, and `rdfs:Datatype` to define custom data types,
- `rdfs:label` and `rdfs:comment`, to provide a human-readable name and description for a resource.

3.4.3 From Vocabularies to Ontologies: the OWL Web Ontology Language

OWL extends the features available in RDFS, allowing the design of complex ontologies and allowing advanced computer reasoning. OWL is divided in three layers of complexity called profiles: OWL Lite, OWL DL (Description Logic), and OWL Full. **OWL Lite** uses all the resources provided by RDFS, with additional features:

- **Classes and properties equality and inequality**: in OWL Lite, `equivalentClass` and `equivalentProperty` indicate synonymous classes and properties. This can be useful, for instance, to make inferences over two different ontologies representing similar domains using differently named classes. Similarly, `sameAs` indicates that two individual resources are the same, and `differentFrom` highlights that two individuals are distinct. `AllDifferent` indicates that all individuals in a set of resources are distinct from one another.

- **Property Characteristics**: a property may be stated to be the inverse of another property using `inverseOf`. In the example from the previous section, the property `hasChild` can be stated as the inverse of a property called `hasParent`. A property can also be stated to be *transitive*, *symmetric*, *functional* or *inverse functional*.

- **Property Characteristics and Restrictions**: restrictions can be imposed on the value that properties can take (`allValuesFrom` and `someValuesFrom`). Restrictions can also be imposed on the number of property values that a class member can hold for a given property (`cardinality`, `minCardinality` and `maxCardinality`). In OWL Lite, cardinality can only take 0 and 1 as a value.

- **Class Intersection**: a given class can be defined as the intersection of two or more classes using `intersectionOf`.

OWL DL and OWL full use all the vocabulary defined for RDFS and OWL Lite, with additional classes allowing features such as the definition of a class by enumeration of all its members (`oneOf`), the restriction of a property to a given individual (`hasValue`), and complex class definitions (disjoint classes, unions, complementary classes, intersection, etc.). Additionally, OWL DL and OWL full allow any non-negative integer value for cardinality (`cardinality`, `minCardinality` and `maxCardinality`). OWL DL and OWL full share the same vocabulary, but OWL DL is subject to some restrictions in the use of this vocabulary. OWL DL requires type separation, which means that a resource cannot be defined as a class and as an individual or property at the same time. This ensures that all conclusions provided by a reasoner from an OWL DL ontology will be computable in a finite time. A detailed description of OWL is provided by McGuinness and van Harmelen (2004).

OWL 2 (W3C OWL Working Group, 2012) is a more recent version of OWL, which has a very similar structure to OWL1 overall. OWL 2 introduces new features such as extended datatypes, extended annotations features and three new profiles: OWL 2 EL used for large ontologies, OWL 2 QL designed for simple ontologies that cover large datasets, and OWL 2 RL for performing complex reasoning in an efficient manner by using a slightly less expressive subset of OWL 2's features.

Automatic Reasoning and Inferences: Deducing Information from Existing Data

In order for automatic reasoners to provide meaningful results, one possible method is to define rules that enable a program to deduce new information from existing data. For instance, the following rule expresses the transitivity of the `rdfs:subClassOf` property defined in RDFS: if a resource `?x` belongs to a class `?y`, and `?y` is a subclass of another class `?z`, then the program concludes that `?x` is also a member of `?z`:

```
{?x rdf:type ?y. ?y rdfs:subClassOf ?z.} => {?x rdf:type ?y}
```

In practice, the program will add a new inferred statement for every matching graph pattern described in the rule. Two languages exist to describe and exchange rules in the semantic web: Rule Interchange Format(RIF), used to enable the transmission of rules from one program to the other across the web, and SWRL(Semantic Web Rule Language), designed to be described rules applied to OWL ontologies.

Tableau-based reasoning can also be used, which involves checking all possible data against a possible conclusion to verify that the proposed conclusion is true given the data provided (Möller and Haarslev, 2009).

Programs that generate inferences from data written in RDFS or OWL are called reasoners. However, in SemVP (the semantic virtual patient system presented in chapter **??** inferences where generated without the help of a reasoner, and using SPARQL in SemVP), due to the overhead involved in running an OWL reasoner and the relative simplicity of the virtual patient model used in SemVP.

RDFS and OWL are languages based on RDF that allow the design of controlled vocabularies and ontologies describing a given domain of knowledge, in order to make meaningful queries and automatic inferences about this domain using an RDFS or OWL reasoner.

The Semantic Web is a group of technologies used to describe, process and combine knowledge on the web. RDF is designed to make statements about any resource with a unique URI, and SPARQL allows the retrieval of RDF data. RDFS and OWL allow the design of vocabularies and ontologies that describe knowledge domains. Rules and reasoners allow conclusions to be automatically drawn from existing semantic data. This research is focused on the Semantic Web because it presents many benefits over existing data formats currently used to design virtual patients.

3.5 Benefits of the Semantic Web over Relational Databases and XML

3.5.1 Benefits of the Semantic Web over Relational Databases

The Semantic Web represents a paradigm shift from the relational database model commonly used in information management systems today. While both technologies are designed to create, combine and consume structured data, relational databases rely on tables in databases to do so, while the semantic web is centered around networks of interconnected nodes. This difference has crucial implications in practice.

In a relational database, pieces of data are grouped together in tables. Each table contains rows, and each row is identified by a unique *key*. Relationships between rows are created by referencing their keys in a new table. This model is efficient and enables fast querying of data, but is not designed to accommodate the easy exchange and combination of data between various data sources or to enable automatic reasoning in the manner described in section 3.4.3.

Structured Query Language (SQL[2]), the ISO standard for querying relational databases, is not designed to accommodate the combination of data from various source in this way, which makes reusing data from one database to another difficult. In addition, it is not possible to query a relational database through HTTP without using intermediary software. By contrast, the Semantic Web is specifically designed to share data across the web, and is standardised in order to achieve this. Every node in RDF has a unique URI, and is therefore accessible from anywhere on the web. RDF data can also be queried directly through HTTP using SPARQL, enabling the retrieval and transfer of data from multiple sources, regardless of the system used for data management. Furthermore, Semantic Web languages are well suited to design complex data. RDF triples represent simple statements, which are interconnected in complex graphs structures, allowing complex and interconnected representations of data. Additionally, RDFS and OWL enable the creation of complex vocabularies and ontologies that describe knowledge in a convenient manner.

3.5.2 Benefits of the Semantic Web over XML: the Example of Medbiquitous XML

XML(Bray et al., 2008) is a language designed to serialize data. This means that XML provides a convenient way to *represent structured information* in order to exchange it between computer programs.

[2] `http://www.iso.org/iso/catalogue_detail.htm?csnumber=4549`

RDF pushes this representation further by providing a *data model* for structured information. RDF, as described in section 3.2, is designed to represent statements in the form of nodes. These statements are not a way to represent data, but rather a representation of the underlying meaning (or *semantics*). RDF is independent from a specific serialization(or representation). Indeed, RDF data can can be represented using XML. Thus, RDF can be considered to be a level "above" XML in its ability to represent information(as illustrated on figure 3.1).

The Medbiquitous virtual patient standard (Ellaway et al., 2010) is an XML specification developed by the Medbiquitous Consortium[3], which aims to be a standard exchange format for virtual cases. The ultimate goal of MVP profiles is to allow the easy transfer of virtual cases from one virtual patient system to another. The MVP standard is encoded in XML documents, associated with existing e-learning formats such as IMS content packaging[4] to catalogue media resources. MVP documents contain a description of virtual patients and virtual cases, divided into four modules to allow maximum flexibility:

1. *Virtual Patient Data (VPD)* represents all the clinically relevant data about a patient, such as questions and answers for medical history, examination and findings, laboratory tests and results, and medical procedures and their outcomes. Most of the VPD data is entered in plain text, with no underlying semantics.

2. *Media Resources (MR)* contain references to the media files associated to the virtual patient: images, animations, videos, audio. Like VPD, MR can be disclosed to the user in response to specific actions. MR are catalogued using IMS content packaging[5].

3. The *Data Availability Model (DAM)* specifies the sequencing in disclosing Virtual Patient Data and Media Resources to the user.

4. The *Activity Model (AM)* describes how the learner will be able to engage with the virtual patient. Various activities from simple observation to decision making are available to the virtual case author. It is possible to design several activity models using the same Virtual Patient Data, Media Resources and Data Availability Model.

[3] `http://www.medbiq.org/index.html`
[4] `http://www.imsglobal.org/content/packaging/`
[5] content packaging specifications: `http://www.imsglobal.org/content/packaging/`

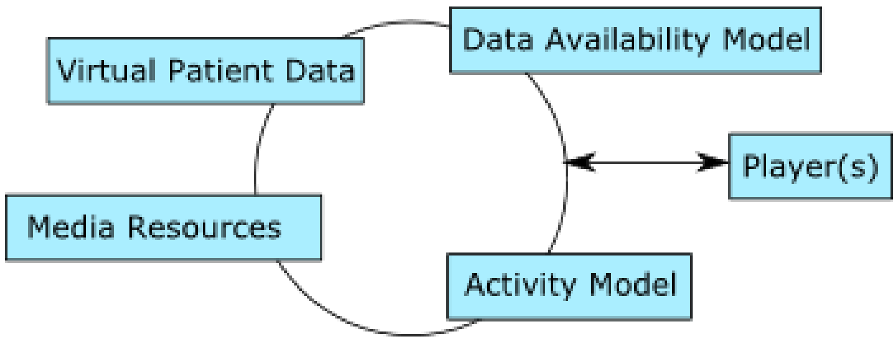


Figure 3.5: The MVP model architecture

The MVP model and its implementation in XML constitute a useful way of structuring and exchanging virtual patients data. Kononowicz et al. (2009) show how MVP allows interoperability between four virtual patients systems: CAMPUS, CASUS, Web-SP and OpenLabyrinth. The export operation from one system to another is automatic. However, importing a profile into a specific system still requires some manual editing, since most of the content in Medbiquitous files is in fact manually authored text with no semantic description, since XML does not provide a data structure to represent the semantics.

The medbiquitous MVP model represents a great potential for virtual patient systems. However, since it is based on XML alone and does not use the semantic features provided by RDF, this model does not allow the generation of automatic feedback using inferences from data representing the student's actions and the patient's condition. As will be demonstrated in chapter 6, a semantic model of virtual patients enables the automated generation of feedback for students, thus increasing the pedagogical value of virtual patients. The semantic web also facilitates the re-use of pre-existing data from various source such as OWL ontologies and medical knowledge bases, for the design of virtual patients.

3.6 Existing Ontologies for Virtual Humans and Biomedical Modelling

The aim of this research is to establish the relevance of Semantic Web technologies for the design of virtual patients, specifically for dynamic feedback generation. This section is an overview of web ontologies presenting a potential use for this purpose.

Biomedical ontologies are a relevant foundation for the design of a Semantic Web-based virtual patient system (see Chapter 6). Formal ontologies and ontology design patterns are also practical tools for the design of high-level concepts and recurrent data modeling problems.

3.6.1 Biomedical Ontologies

Biomedical ontologies provide data models to describe various fields in the biomedical domain: anatomy and physiology (Golbreich et al., 2006; Grenon et al., 2004), genetics and biology (Qu et al., 2009). Biomedical ontologies are built to serve many purposes: research in biology, pharmaceutical, genetics, molecular biology, etc.

They vary greatly in level of granularity and in the breadth of the domain they cover. Several ontologies and groups of biomedical and genetics ontologies present a potential for this research:

- Unified Medical Language System (UMLS) contains a variety of ontologies, including the Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003) and the International Classification of Diseases devised by the World Health Organization. UMLS also includes the SNOMED terminology (Systematised Nomenclature of Medicine Clinical Terms), which will be used by the NHS as a standard to facilitate communication between health-care professionals[6]. Unfortunately, most of the data from the UMLS is proprietary and is not available as linked data or in the form of OWL files. The FMA, however, is available in OWL.
- The Open Biomedical Ontologies (OBO) consortium is pursuing the integration of many different biomedical ontologies in a coherent model, in order to facilitate their integration(Smith et al., 2007). The OBO foundry contains ontologies covering domains such as cellular components, phenotypical qualities, protein structures, anatomy, etc. However, OBO ontologies do not cover medical knowledge such as medical procedures, conditions, symptoms, etc.
- *OpenGalen*[7] is an ontology designed with the same goals as UMLS, and available under an open source licence (Rector et al., 2003). OpenGalen is extensive, and provides data on anatomy, drugs, genetics, social factors, and various elements specific to many medical specialities such as gastrointestinal, nutrition, endocrine, oncology, and many more. OpenGalen is an extensive group of ontologies and is the most appropriate option for the design of a semantic virtual patient model. Indeed, OpenGalen is an ideal combination of high-level classes describing both everyday medical interactions and low-level biological concepts. It has been chosen for the design of SemVP, the Semantic Web-based virtual patient system developed for the purpose of this research (see Chapter 6).

[6] `http://www.connectingforhealth.nhs.uk/systemsandservices/data/snomed`, "Connecting for health-NHS"

[7] `http://www.opengalen.org/index.html`

3.6.2 Formal Ontologies and Ontology Design Patterns

Formal ontologies are designed to cover very generic concepts and provide building blocks for other, more specific ontologies. By doing so, they help alleviate some semantic ambiguities, and provide logically sound definitions to draw from when designing more domain-specific ontologies. The WonderWeb ontology[8] is an example of such and ontology (Masolo et al., 2003). Cyc (Matuszek et al., 2006), although not strictly a formal ontology, features many formal concepts used throughout the ontology. Cyc and its open source subset OpenCyc contains knowledge regarding a large variety of domains.

Unfortunately, general formal ontologies tend to become very complex, very large and thus extremely difficult to maintain. To enable the modelling of generic concepts in a more flexible way, another approach is the use of ontology design patterns, inspired by object-oriented programming design patterns (Gamma et al., 1995). Ontology design patterns are very small ontologies designed to solve one specific modelling problem (Gangemi, 2005). Therefore, when confronted to a specific ontology design problem, it is possible to integrate the pattern to a pre-existing ontology without loading a large formal ontology.

A repository of ontology design patterns ready to be used can be found at `http://ontologydesignpatterns.org/wiki/Main_Page`. Design patterns were used in the design of SemVP, as described in Chapter 6.

The Semantic Web is a group of technologies designed for the representation and exchange of data across the web. Vocabularies and ontologies can be designed in RDFS and OWL, allowing automatic reasoning over web data. These features, in addition to existing ontologies and specifications, constitute the foundations of SemVP, a Semantic Web-based Virtual Patient system presented in Chapter 6.

[8] `http://wonderweb.semanticweb.org/`

Chapter 4

Feedback in Virtual Patients

This chapter proposes a definition of feedback in education, from the literature on clinical reasoning, in particular looking at the differences in reasoning strategies between students and experienced clinicians. Schema theory was found to be a suitable framework to understand how clinical expertise is acquired. Feedback is examined as a way to help students to elaborate schemes that will help them make appropriate decisions quickly in clinical settings.
Using this analysis, a classification of seven different types of feedback for self-assessment virtual patient systems was compiled. In the light of this analysis, the semantic web has many benefits for the dynamic generation of feedback in virtual patients.

4.1 Definition of Feedback

Before defining the most useful types of feedback in virtual patients, a clear definition of feedback, as understood in education, needs to be established. The nature and purpose of feedback in medical education is not always clear, which sometimes lead to serious pedagogical issues stemming from differences in view about feedback between students and teachers (Sender Liberman et al., 2005; McIlwrick et al., 2006; Gil et al., 1984). To clarify this issue, it is useful to consider feedback in the broader context of student assessment, in particular (for the purposes of this research) formative assessment. Assessment can be understood as a set of activities performed by students and teachers to provide information to be used as feedback, in order to modify the teaching and learning activities they engage in (Black and Wiliam, 2006). This definition puts feedback at the heart of the assessment process. Wood (2011) provides a framework to explain the positive outcomes of formative assessment, such as self-regulation and independent initiative, autonomous learning skills and resourcefulness.

Van de Ridder et al. (2008) explicitly set out to formulate a clearer definition of feedback in clinical education, and reviewed the relevant literature in social sciences, medical education and other fields. The result of this review is the following operational definition: *"Specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance"*. This definition fits perfectly to the use of virtual patients as an self-assessment process, were students perform an activity (attempting to determine the patient's diagnosis and management) which is "observed" using the virtual patient system, and feedback is provided to them, comparing their clinical process to one used by a more experienced clinicians (which, in this situation, is usually the virtual case author). The goal of this activity is to improve students' clinical reasoning skills, helping them reach a standard, which in this situation is the case author's reasoning abilities. In other words, virtual patients as self-assessment tools are designed to help students solve clinical cases increasingly more like an experienced clinician would, through an observation of their choices within a virtual case, and (crucially) subsequent delivery of feedback about these choices.

Van de Ridder et al. (2008) also identify three underlying concepts defining feedback: feedback as information, feedback as "reaction", and feedback as a "cycle". At its most basic level, feedback can simply be thought of as information provided to the learner about the performance observed (Black and Wiliam, 2006). This definition is focused on the feedback message itself, but considering feedback as a "reaction" (typically, the teacher's reaction to the student's actions) takes into account the actors involved in feedback delivery and reception. Feedback as "cycle" describes an ongoing process involving recurrent exchanges of information between the learner and the teacher (or virtual patient system). In this process, the common goal for the learner and the teacher is to reduce the gap between the student's skills and the teacher's ability (or a standard level of ability defined by the virtual case author) over time.

In the context of virtual patients, feedback on each virtual case can be designed as generic information explaining the "correct" course of action (e.i. the choices recommended by the case's author or authors, defined as a standard to reach) and the rationale behind it. It can also be regarded as a "reaction" to the student's actions in the virtual patient system. Feedback as a "cycle" can also be achieved in virtual patients, by providing information to students at each step of their progression in the case (sometimes with the option for students to change their choice after receiving feedback). The study of several similar virtual cases over time, with feedback at the end of each case, can also be considered as feedback as a "cycle", since the student can increasingly improve on each case using feedback provided from the previous cases.

This research is mainly focused on feedback as a personalised "reaction" to students' choices, since it is focused on automatic feedback generation. Evaluating feedback delivered as a "cycle" would involve testing multiple ways of delivering feedback over long periods of time, which is outside the scope of this research, primarily concerned with the use of semantic web technology to generate useful information to students, regardless of the time and place of delivery. Investigating the pedagogical impact of various feedback delivery methods to optimise feedback as “cycle” is a valuable future direction for this research, as described in Chapter 9.

Various aspects of feedback can contribute to its effectiveness or “strength”. Four aspects of effective feedback can be extrapolated from the definition proposed by Van de Ridder et al. (2008):

- *Explicit Goal*: the objective or standard that the feedback is supposed to help the student reach has to be clearly understood by both students and teacher,
- *Specificity*: feedback needs to refer to clear aspects of the student’s performance,
- *Observable Actions and Skills*: in order to be specific, feedback has to refer to clearly observable (and observed) actions or skills so as to provide information or judgment about these actions or skills,
- *Personalised*: each student’s actions are unique, in that they make different mistakes and need to alter their behavior in different ways in order to reach the intended standards. Therefore, each student requires feedback that is specific to their own actions and mistakes. This is an important aspect of this research, since creating personalised feedback is a time consuming task for virtual case authors, which can be facilitated by semantic web technologies.

Other aspects of feedback can contribute to the strength of in-person feedback, such as re-observing the student after an initial feedback, delivering feedback in a non-judgemental fashion and leaving time for students to react (Sender Liberman et al., 2005). However, these features mostly apply to teachers’ interpersonal skills, and are difficult to implement in a virtual patient system. An evaluation of the emotional aspect of feedback delivery is outside the scope of this study, but would constitute a valuable future research direction (see Chapter 9).

4.2 Using Feedback to Improve Clinical Reasoning in Virtual Patients

In the literature, feedback is mostly referred to as an exchange of information between two people, the student and the teacher. In the context of virtual patients for self-assessment, the virtual patient system can be envisioned as an “intermediary device” between the user (or student) and the case author (or teacher), facilitating an asynchronous “conversation” between both actors of the assessment process. This has practical benefits, such as the reusability of the system without the need for a clinician to be present to provide feedback. It also has pedagogical benefits, since virtual cases are standardised by definition and allow students to practice on cases which fit the curriculum and their learning needs, and to receive consistent feedback every time. However, to obtain these positive outcomes from virtual patients, strong feedback needs to be provided by virtual patient systems. In the light of the definition provided above, this means that, to be effective, virtual patient systems need to provide specific and personalised feedback about observed actions within the virtual case. To be effective, the feedback also needs to be designed with the goal of improving students' clinical reasoning. This is especially challenging in a situation where the teacher is not present to deliver feedback in person, but needs to provide information for a large group of different students, and thus take into account all the most common mistakes and the good practices to remember for each case. As a result, a clearer understanding of clinical reasoning in general is needed in order to design and generate useful feedback. In particular, the differences between the clinical reasoning strategies used by students and those used by clinicians need to be identified and understood.

Clinical reasoning is an important aspect of medicine, and a crucial but difficult skill to teach medical students. Researchers in medical education and cognitive psychology seek to understand the processes through which clinicians reach an appropriate diagnosis and subsequently manage the patient's condition. This research has key implications in medical education. An understanding of how and why medical students' thought processes differ from those of experienced clinicians (experts) is a solid foundation for the design of effective feedback that facilitates students' transition from their initial level to a higher standard of clinical reasoning.

Norman (2005) identifies three consecutive trends in the history of research in clinical reasoning, spanning over the last 30 years:

1. Attempts to understand clinical reasoning as a generic skill, independent of specific medical knowledge,
2. Models of clinical reasoning based on memory and knowledge,
3. Research focusing on how mental representations of knowledge such as scripts and schemes are used by students and experts to reach a diagnosis.

Elstein et al. (2002) argue that modelling the clinical reasoning process solely as a generic hypothetico-deductive process leads to an incomplete understanding of expertise in clinical reasoning. Even though the reasoning process can be analytically separated from the domain knowledge, in practice this model fails to take into account the effect of clinical experience on the accuracy and speed of the diagnosis process. Consequently, hypothetico-deductive reasoning does not explain why some students struggle to generate appropriate hypotheses from the clinical data they gather, and why clinicians generate few specific hypotheses that they verify quickly and efficiently. Thus, the classic Bayesian reasoning method sometimes conflicts with clinical experience. Elstein et al. (2002) show that experienced clinicians often use alternative strategies such as pattern recognition and scheme-inductive reasoning.

Pattern recognition is the use of experts' extensive experience in clinical practice, quickly generating diagnosis hypotheses using memory from previously encountered symptoms or combination of symptoms on other patients. One could assume that using this strategy alone can lead clinicians to make premature and erroneous diagnoses (also called "faulty triggering", see Kassirer and Kopelman (1989)). However, most studies examined for this feature show that experts generally make less diagnosis mistakes than students, even though they reach the diagnosis more quickly, using fewer initial hypotheses and asking fewer questions. Thus, it appears the models based on memory and knowledge retrieval alone are insufficient to fully represent the diagnosis process. Other factors must be involved in the acquisition of clinical expertise.

Biomedical knowledge alone is not sufficient, and it appears from the literature that the way knowledge is structured and connected plays a central role in clinical expertise. Rikers et al. (2000) show that expert clinicians, through experience, encapsulate medical knowledge in structures allowing them to access and use information quickly in a clinical situation. This way of processing information is used by clinicians even outside their specific domain of expertise, and allows them to reach appropriate conclusions faster than students regardless of the medical discipline. Mandin et al. (1997) confirm that knowledge organisation is a crucial feature of medical expertise, and that experts do indeed organise knowledge in the form of schemes, which include both conceptual and procedural knowledge.
Schmidt and Boshuizen (1993) describe the evolution of medical competency towards expertise in three phases:

1. Acquisition of causal knowledge about diseases and their consequences (the "basics" of medicine), typically in the first years of study,
2. Elaboration of narrative structures called illness scripts through experience with real cases (starting with students' first clinical assignments),
3. Use of experience from previous cases combined with encapsulated biomedical knowledge in the diagnosis of new cases.

Boshuizen et al. (1995) confirm this theory by arguing that advanced students are indeed very knowledgeable about conditions in patients and the environment, but are unable to apply this knowledge adequately in clinical reasoning. To acquire this ability, students need to integrate and structure their existing knowledge in more efficient mental structures, grouped around conceptual clusters. The challenge for feedback in virtual patients, therefore, lies in enabling a transition from simple causal knowledge to encapsulated and actionable knowledge allowing fast and reliable diagnoses.

Illness scripts are also mentioned by Charlin et al. (2007), who used script theory as an assessment tool for clinicians. Scripts are goal-oriented narrative structures that help practitioners give meaning to new situations. In clinical practice, this means that certain combination of symptoms, signs and contextual information about patients will lead practitioners to make certain inferences and to perform certain actions. For instance, if a baby presents with a fever associated with general weakness and fatigue, most students will think about meningitis immediately and check symptoms such as photophobia, rash, etc. to confirm this diagnosis. Scripts can be seen as procedural knowledge, and allow advanced students and young doctors to make quick decisions. Scripts also include expected values and normal ranges for various parameters, allowing clinicians to verify their diagnosis hypothesis. In the example of the baby with a fever, students and young clinicians will measure temperature and examine the infant to check for fever or rashes, in order to confirm or rule out meningitis.

To acquire a greater level of expertise and a higher level of diagnosis reliability, procedural knowledge structured in scripts has to be combined with conceptual knowledge, allowing practitioners to deal with “fringe” cases or unexpected results more efficiently. Scheme-inductive reasoning uses information from previous cases in an elaborate manner. Schema theory describes mental structures that combine procedural and conceptual knowledge (Van Gog et al., 2004; Gauthier et al., 2008; Marshall, 1995). Van Gog et al. (2004) also highlight the presence of both strategic and principled information in schemes. This means that when concepts are stored using schemes, they are not just retained as abstract ideas, but are usable to make decisions. This explains the findings made by Rikers et al. (2000), showing that experts can retrieve and use biomedical concepts faster than students, since experts tend to structure their knowledge in schemes. Additionally, schemes contain deeply structured information, preventing faulty triggering errors occurring when simple pattern recognition is used. A scheme is a memory structure allowing a practitioner to recognise common patterns, but also to elaborate from similar experiences and to make inferences based in this preexisting knowledge, in order to plan a course of action.

Scripts and schemes are typically acquired and altered through experience. They provide specific knowledge from memory of specific events, but they are also flexible enough to allow adaption to a new problem or a new presentation of an existing problem. Indeed, in scheme-inductive reasoning, memories of previous cases are used to reinforce or modify complex memory structures. This explains why scheme-inductive reasoning can dramatically increase diagnosis success compared to hypothetico-deductive reasoning (Coderre et al., 2003).

4.2.1 The Role of Feedback in Reflective Practice

Schema in clinical reasoning emerge from the encapsulation of biomedical knowledge, associated with the acquisition of illness scripts from clinical experience. Both knowledge encapsulation and the formation of illness scripts are facilitated by reflection and self-assessment about encountered cases, especially for students transitioning from early experience to clinical expertise. For students, feedback about their performance on cases they work on is a central part of this reflective practice. As a result, feedback is at the heart of clinical reasoning skills acquisition and improvement.

Marshall (1995) emphasises the impact of repetition on schemes building. The practice of several similar scenarios enables the elaboration of schemes. Indeed, a scheme can be elaborated from a single situation, but is strengthened and deepened when it emerges from common patterns on which students can elaborate to make future decisions, using encapsulated biomedical knowledge as a tool for verification and causal explanation. The passive study of multiple cases is not sufficient to facilitate expertise acquisition. A reflective assessment of each case, with feedback delivered through self-evaluation, peer collaboration and tutoring, is necessary.
In medical education, working through multiple cases around the same topic can help students alter and reinforce their diagnostic schemes. For instance, by studying multiple patients presenting a headache, students reinforce their understanding of the causes underlying this symptom and learn to apply this knowledge appropriately in context. This is a typical example of simultaneous knowledge encapsulation and illness script elaboration, leading to schemes elaboration. One particular scheme could be centered around neurological conditions (prompting students to investigate the patient's neurological history and symptoms), and another could relate to the circulatory system (migraines). Each scheme is triggered depending on the case's context: nature and precise location of the pain, patient's history, patient's age, etc.. As an example, to facilitate the formation of these schemes, feedback should remind the student that headaches can be symptoms of neurological conditions in some situations, and can also be symptoms of circulatory problems in other cases. The appropriate actions to perform to check for each type of condition also have to be specified in the feedback. If students neglect observations that should be performed to confirm or rule out a likely condition, feedback should also remind the student of this oversight, and explain why such observation should be performed in this case. Such feedback will reinforce student's schemes and trigger the appropriate actions when similar situations arise in the future.

Another way to use multiple cases for schemes elaboration is to practice several cases in the same medical specialty, or focused on the same body system. This enables students to practice similar clinical assessment procedure in various contexts and to understand the commonalities and differences between each case. The resulting schemes help students adapt to new situations, while enabling a rapid recognition of common features and differentiating between the most probable causes of common symptoms.

Gauthier et al. (2008) argue that experts decision maps representing experts' schemes can be used to design feedback which helps students to develop their expertise in clinical reasoning. For example, negative knowledge (i.e. knowledge of what *not* to do) helps students understand actions to avoid in real clinical situations (Gartmeier et al., 2008).

Elstein et al. (2002) propose that problem-based learning (PBL) supports the learning of hypothetico-deductive clinical reasoning. It can be proposed that PBL also supports schemes elaboration through reflection on multiple cases. This reflective practice is supported in PBL by student collaboration and exchange of knowledge, self-assessment and tutor feedback. Feedback from peers and teachers is indeed a key component of PBL (Albanese, 2011).

Feedback regarding students choices of interaction with virtual cases helps them elaborate and structure their medical knowledge into schemes. Thus, feedback is a key feature in virtual patients. As a consequence, the semantic web-based system created for this research was designed and evaluated with a strong emphasis on feedback.

4.3 A Proposed Classification of Feedback Types

It emerges from the literature cited above that feedback is best used to encourage students to reflect on their actions, in order to:

- Understand which choices of questions, examinations or lab tests are most appropriate, which are less appropriate, and which can have harmful consequences on a patient,
- Understand why a choice is appropriate or not for each patient.

Kassirer and Kopelman (1989) and Friedman et al. (1998) clarified the nature of medical errors by elaborating classifications based on a four-step process of problem solving rooted in cognitive science (see Langley et al. (1987) and Pólya (1957)):

1. Develop a cognitive representation of the problem and trigger hypotheses,
2. Make a plan (determine how to verify or invalidate clinical hypotheses),
3. Carry out the plan (in clinical practice, this means gathering and processing information through questions, clinical examinations or laboratory test),
4. Look back at the results and verify the hypotheses generated in step 1.

It is safe to assume that a similar process would apply to patient management, but this can be considered as a second phase in the process. Kassirer and Kopelman (1989) and Friedman et al. (1998) focus only on reaching a diagnosis, and this research is also limited to diagnosis, since it is mainly concerned with students in year 3 (starting clinical assignments). Patient management is the main learning outcome at the Southampton School of Medicine starting in year 4.

Using this process as a basic framework, the following types of medical errors can be extracted from Friedman et al. (1998) and Kassirer and Kopelman (1989):

- *Faulty Hypothesis Triggering* resulting in accurate findings used incorrectly and fabricated findings,
- *Faulty Context* (failure to formulate an assessment plan adapted to the case's situation),
- *Faulty Information Processing* resulting in non-discriminatory findings used to support diagnosis, over-reliance on axioms and faulty data interpretation,
- *Faulty verification.*

A proposed classification of feedback types is presented in this chapter. This classification synthesises several studies regarding students diagnosis mistakes and the difference between students diagnosis process and experts diagnosis process (Friedman et al., 1998; Kassirer and Kopelman, 1989; Gauthier et al., 2008; Gartmeier et al., 2008; Charlin et al., 2007; Van Gog et al., 2004; Marshall, 1995). These studies all confirm the role of scheme-inductive reasoning for clinical experts.
It is expected that such a classification will not only help to design a useful feedback generator for virtual patients, but also improve the general understanding of medical students' pedagogical needs in clinical reasoning.
Based on the literature on clinical reasoning and cognition described above, three criteria have been used to elaborate feedback types that are likely to facilitate schemes elaboration for students:

- *Feedback has to be focused on each student's actions*: Since feedback is meant to bridge the gap between each student and an given standard of expertise, feedback is more effective when it starts from each student's knowledge and skills. The simplest way to establish a student's level of expertise is to observe the student's actions on a given case. As described in Van de Ridder et al. (2008), feedback is a "reaction" to observed actions performed by a student.
- *Feedback has to emphasise the appropriate choices made by students as well as highlighting the irrelevant or even harmful choices, and explain the underlying*

reasons in each case. To elaborate appropriate and useful schemes, students need to understand their successes as well as their mistakes, in order to reproduce correct choices in the right context and to change their behaviour when needed.

- *Feedback has to communicate the standard to reach clearly*: Van de Ridder et al. (2008)'s definition of feedback specifies that feedback starts with standard to reach for students and agreement between the students and the teacher (or case author) about this standard is paramount for effective feedback. In the context of virtual patients, the standard to reach is defined by the case author, often in the form of a sequence of appropriate actions to perform in each case.

Using these criteria in the context of feedback in virtual cases, the following classification is proposed:

- *Feedback Type 1*: "A list of interview questions, examinations and tests that the student should *NOT* have chosen, and the justification (not appropriate, irrelevant, redundant, etc.)".
  This type of feedback helps students reflect on faulty hypothesis triggering (Friedman et al., 1998; Kassirer and Kopelman, 1989), by pointing out that the student investigated an unlikely condition, sometimes at the expense of more probable diagnosis hypotheses.

- *Feedback Type 2*: "A list of the interview questions and examinations that students *should* have chosen, and the justification (type of disease to consider, related symptom to check, possible conditions to rule out, etc.)".
  This helps students reflect on faulty triggering, by identifying questions and examinations related to a hypothesis they might have neglected. It can also help students identify faulty verification.

- *Feedback Type 3*: "If the diagnosis is wrong, feedback telling the student if the chosen diagnosis is still coherent with the results of the chosen interview questions and examinations".
  This type of feedback helps students reflect on information processing and on hypothesis triggering.

- *Feedback Type 4*: "Feedback about the order in which the student performed specific actions".
  This type of feedback deals with the logical flow of diagnosis reasoning that an expert (the case author) would follow. This logical flow can be represented using expert decision map (Gauthier et al., 2008). This type of feedback is essentially a comparison between the student's and the case author's decision maps. It also helps students reflect on faulty information processing.

- *Feedback Type 5*: “A sequence of the “ideal” history taking and examination process that an expert would use, with the rational for each step”.
  This feedback is very much related to feedback type 4, and helps students understand the process of diagnosis for each case by example. The experts decision map is very explicit in this type of feedback, which clearly communicates the standard the student needs to reach.

- *Feedback Type 6*: “If the student chooses an inadequate action, a narrative description of the consequences on the patient, if applicable”.
  This can apply to faulty context (failure to identify pre-existing condition, allergies, etc.), faulty hypothesis testing, and faulty verification. This type of feedback can sometimes be appropriate for history taking and diagnosis, but it is most appropriate for feedback on patient management.

- *Feedback Type 7*: “A list of all diagnoses the student should have tested and ruled out, given the initial presentation of the patient”.
  Friedman et al. (1998) and Mandin et al. (1997) suggest that experts tend to generate a smaller number of diagnosis hypotheses, using their experience to focus on the most likely possibilities. This type of feedback could help students, after seeing multiple examples and suggestions of hypotheses, to generate more accurate hypotheses in the differential diagnosis when confronted with various types of patient, thus reducing the frequency of faulty triggering and faulty hypothesis testing. It also helps students to identify the relevant conditions to rule out.

4.4 Benefits of a Semantic Web Model for Dynamic Feedback Generation

Virtual cases authors have to design feedback with care when creating a virtual case. Feedback is indeed a key component of virtual cases, enabling students to elaborate diagnostic schemes to be used in clinical practice.
In most current virtual patient systems, feedback is written manually, and feedback delivery is organised using linear or branching structures. However, this approach has several limitations

Firstly, writing feedback in this manner is a time consuming task; virtual patient authors have to write feedback for every possible action that a student could select, and every possible consequence of each action. On the other hand, semantic representations (RDF resources) for each action can be linked to the conditions they help confirm or rule out, and this information can be re-used on multiple virtual cases to generate feedback for each action. Additionally, each resource representing an action can be linked to potential consequences with parameters to determine the context in which these consequences can arise. This would also constitute reusable data that case author can harness to design new cases.

Secondly, conventional “static” feedback is unique to each case, and the information contained in the feedback cannot usually be re-used in other contexts. Even a simple translation from one language to another practically amounts the redesign of a whole case. Current specifications (such as Medbiquitous, see Section 3.5.2) are helpful in terms of system-level interoperability, but they do not provide the necessary semantic level to describe the patient’s condition, the available observations and treatments, or the student’s actions in a meaningful and reusable format. This means that, in practice, transferring data from one Medbiquitous-compliant virtual patient system to another without re-editing case content is not possible, despite the design of efficient Medbiquitous conformance test (Kononowicz et al., 2009).
Semantic web technologies can describe a patient’s condition and medical history, the available actions (questions, examinations, and lab tests), and each student’s action within the system in a structured and unambiguous knowledge base that can be exported from one system to another. Using this data, automatic and personalised feedback can be generated in multiple languages and in multiple formats (text, video, still pictures, etc.). Additionally, patient data can be transfered and adapted to create multiple patients with similar features, which would support the elaboration of feedback as “cycle” and facilitate schemes elaboration.

Chapter 5

Preliminary Study

Prior to the design of a semantic web-based virtual patient system called SemVP (see Chapter 6), a preliminary study was conducted. A survey among students and clinicians from the University of Southampton Faculty of Medicine, as well as qualitative interviews with a small group of students, were conducted. The methodology and the results of this study are presented in this chapter

5.1 Survey Research Questions and Methodology

An online survey was conducted among a sample of 16 teaching clinicians and 51 students in year 3 (24 students), 4 (16 students) and 5 (10 students) from the School of Medicine in Southampton. The URL of the survey was sent by email to all students in year 3, 4 and 5, and to the clinicians who are involved in teaching students in those year groups. The survey was available to students and clinicians for a period of 3 months.

The survey was designed to answer the following research questions, and the results were used to guide the design of SemVP and its underlying semantic data model:

- Q1: Which type of virtual patient interaction do students and clinicians have experience with (linear, branching or exploratory cases)? What are their opinions about the cases they used or designed in the past, in particular feedback provided in these cases?
- Q2: What type(s) of virtual patient interaction do students and clinicians find most appropriate for self-study in clinical years (year 3 to 5)? A very simplified model of interaction was submitted to students and clinicians for evaluation.
- Q3: Which aspects of clinical reasoning do students struggle with most?

- Q4: Which type(s) of feedback from the classification presented in Chapter 4 are considered most useful by students, and which ones are considered most useful by clinicians? Is there any other type of feedback that students or clinicians find useful?

Students' opinions about each feedback type do not necessarily reflect their actual pedagogical needs accurately, and are not used as definitive criteria for the design of feedback in SemVP. However, the results of the survey provide useful information about how students perceive their own weaknesses, which is a relevant starting point to design a learning experience that will resonate with students' expectations and current level of expertise. Additionally, feedback types were described in the survey, not as abstract requirements, but as concrete descriptions related to real clinical situations. These descriptions were designed to relate to students' past experience of clinical feedback, which facilitates a recall of past difficulties in clinical situations and is expected to yield more accurate results. Finally, the answers given by clinicians provided a relevant external perspective on the student's opinions.

The survey was designed using the iSurvey system provided by the University of Southampton[1]. It was divided in four sections, which are described below:

1. *Student's or clinician's personal information.*
   This includes the year group, medical speciality (for clinicians) and, optionally, the participant's email (to enter a draw for a £30 Amazon voucher).
2. *Previous Experience with Virtual Patients.*
   This section contains questions regarding previous experiences with virtual patients, such as:
   - Have you ever used a virtual patient in the past?
   - What kind of virtual patient have you used in the past (Linear, Branching, Exploratory, other)?

   Clinicians were given two more questions related to the authoring of virtual patients:
   - Have you ever authored or contributed to a virtual patient in the past?
   - What aspect of the virtual patient design and creation process did you find most difficult or frustrating?

   Participants were then invited to rate the three following statements using 6-steps Likert scales ranging from "Strongly disagree" to "Strongly agree":
   - "The feedback provided by virtual patient systems I used in the past was useful.",

[1] `https://www.isurvey.soton.ac.uk`

- "The feedback provided by virtual patient systems I used in the past was relevant to students' learning needs (or *to my learning needs*) at the time.",
- "Self-study virtual patients would be a useful tool to improve decision-making and diagnosis skills during students' clinical training.".

The scales have no midpoint, in order to obtain a clear answer from participants, in one direction or the other. However, the limitation of this design is acknowledged: Likert scale without midpoint may induce a bias in participants' responses by preventing them from selecting a neutral opinion. To avoid such a bias, Likert scales with 5 steps have been used for the final study of this research (see Chapter 7).
To analyse the results, Likert scales were divided into three categories: the first category contains "Strongly Disagree" and "Disagree", and is considered to be a negative opinion. The second category includes "Somewhat disagree" and "Somewhat agree", and is considered a neutral response. Answers including "Agree" and "Strongly Agree" are considered to be positive opinions.

3. *Expectations and Requirements for a New Virtual Patient System.*
This section is the most important one of the survey, since it was used to guide the design of SemVP's model directly. The first question in this section is *"In your opinion, which aspects of the diagnosis process do students find most difficult?"*. The provided options are:

- "Knowing which conditions to test for given the patient's initial presentation",
- "Knowing the relevant history questions to ask and the relevant examinations to perform given the patients initial presentation",
- "Interpreting the information obtained through interview and examination",
- "Adjusting the differential diagnosis using the patients answer to each question and the result of each examination",
- "Other" (participants could specify in plain text).

The second question in this section is *"In your opinion, what kind of interaction would be most useful to students in a self-learning and self-assessment virtual patient system?"*. The options prsented are:

- "Being guided step by step throughout the case with questions and quizzes with feedback on each question" (linear case),
- "Having a limited number of choices and seeing the consequences of each choice until the case ends in success or failure" (branching cases),
- "Being able to make as many decisions as possible unguided and obtaining a global feedback on each choice made at the end of the case" (exploratory case),

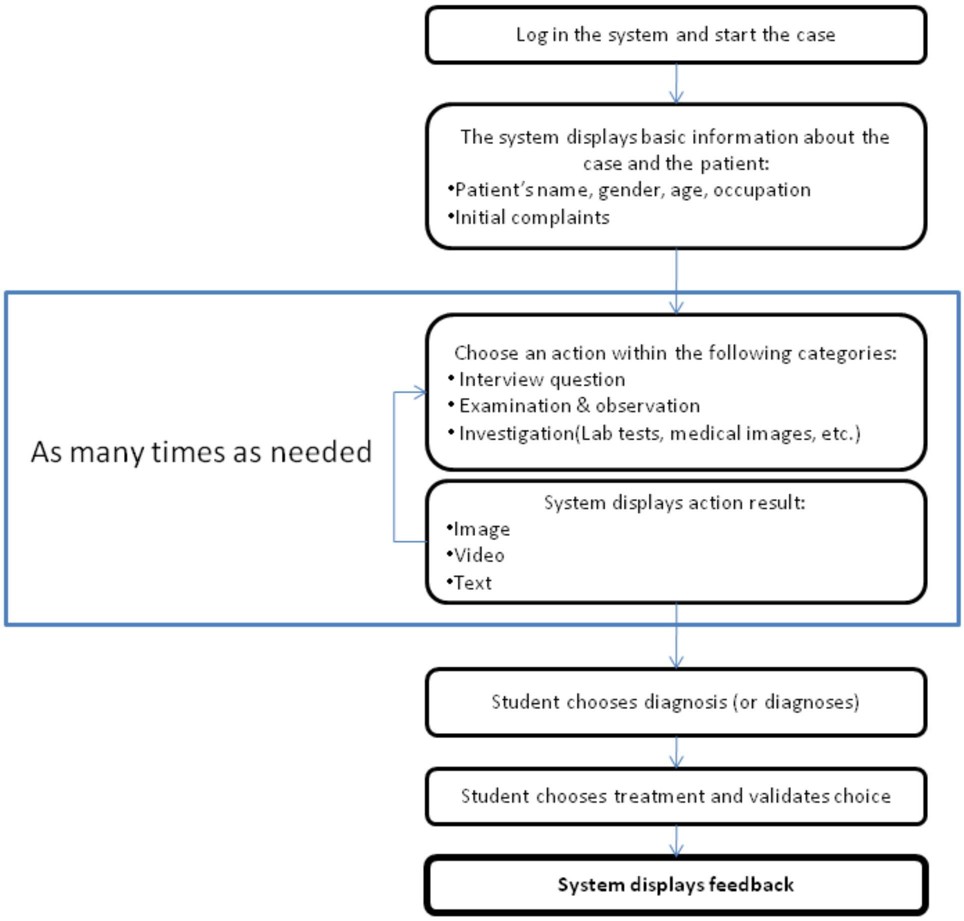


Figure 5.1: Simplified model of virtual patient interaction, submitted to students and clinicians for evaluation.

- "Other" (participants could specify in plain text).

Feedback Types Classification

The last part of this section allows students and clinicians to rate the seven different feedback types described in Section 4.3. Each type of feedback was rated on a 6-steps scale ranging from "Completely Useless" to "Very Useful".

4. *Proposed Virtual Patient System Interaction*
This section proposes a very simplified model of a students interactions with the virtual patients, described in figure 5.1. The aim of this section was to determine if the model of interaction envisioned for SemVP was aligned with how participating students and clinicians think about clinical cases, which contributed to answer research question 2 regarding the most relevant interaction design for self-study virtual patients.

   The two following statements are given to participants to rate using a Likert scale ranging from "Strongly Disagree" to "Strongly Agree":

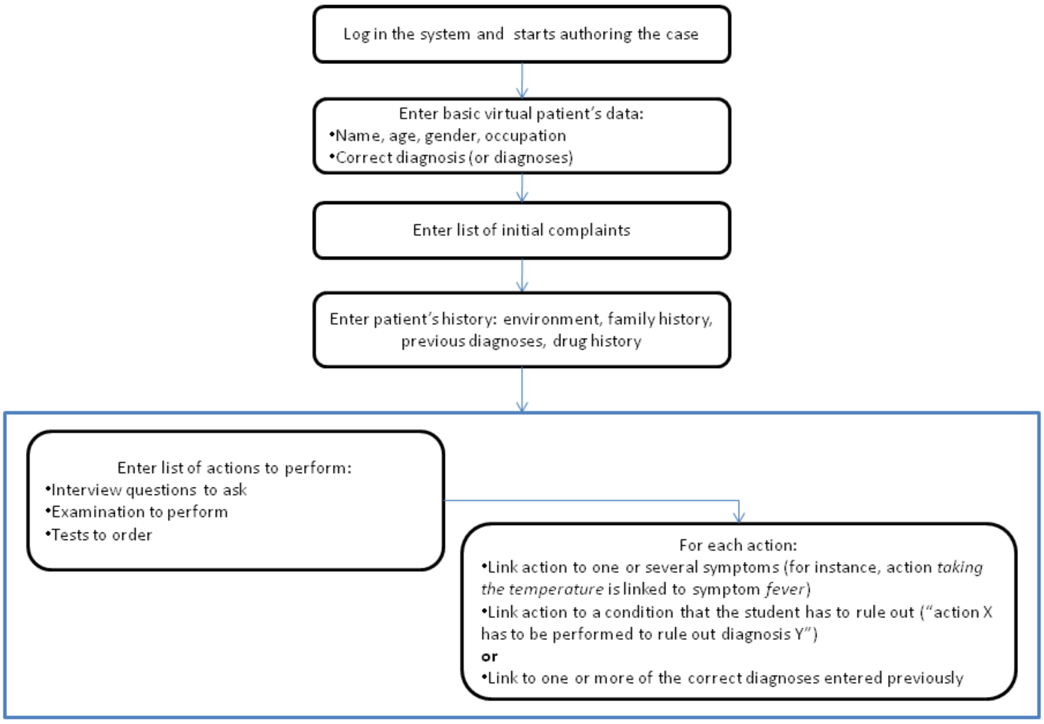


Figure 5.2: Simplified model of virtual patient authoring, submitted to clinicians for evaluation.

- "This process corresponds to the way I think about a clinical case."
- "This process can provide all the information students need to make an accurate basic diagnosis."

The same questions were asked to clinicians regarding a simplified diagram representing the virtual patient authoring process (see figure 5.2).

5.2 Survey Results

Previous Experience and General Opinions about Virtual Patients

Students and clinicians have a different experience of using or authoring virtual patients as teaching and learning tools. Fourty-one participating students (82%) have had experience using virtual patients, while only 6 out of 16 clinicians had experience using a virtual patient, and only 2 had authored a virtual patient before. The type of case most used by students was the linear virtual patient (36 students - 88% - have used this type of virtual patient in the past). Indeed, the virtual patients used as part of the curriculum in year 1 and 2 at the Faculty of Medicine are linear cases. Twenty-four students who used virtual patients in the past (60%) agreed or strongly agreed with the statements "The feedback provided by virtual patient systems I used in the past was useful" and 62% agreed or strongly agreed to the statement: "The feedback provided by virtual patient systems I used in the past was relevant to my needs". Only 5 clinicians answered these questions, but 4 of them agreed with both statements, while the remaining participant was neutral. To the statement "Self-study virtual patients would be a useful tool to improve my decision-making and diagnosis skills during my clinical training", 30 students (60%) either agreed or strongly agreed, which suggests a real demand for virtual patients as practice tools among students.

There was a general consensus that virtual patients are relevant and useful to students in their study. Students have more experience of virtual patients than clinicians due to the deployment of virtual patient systems within the University of Southampton in the past few years.

Expectations and Requirements for a New Virtual Patient System

To the question: "in your opinion, which aspects of the diagnosis process do you find most difficult?", 30 students (62%) selected *"adjusting the differential diagnosis"*. However, only 7 clinicians also considered this aspect to be difficult. Most clinicians (10 out of 16) considered that *interpreting the information obtained through interview and examination* is an aspect of the diagnosis that student struggle with most. Hwever, only 18 students (35%) also considered this aspect difficult. Twenty-six students (60%) selected *"Knowing which condition to test for, given the patient's presentation"*, while only one clinician considered this to be a difficult aspect of diagnosis. This is an important difference (albeit not statistically significant), one that could confirm the experts' ability to generate focused diagnosis hypotheses using schemes elaborated through experience, while students struggle with this aspect of the process (see Chapter 4). This is consistent with findings from Friedman et al. (1998), Kassirer and Kopelman (1989) and Mandin et al. (1997), and shows the need for virtual patients that help students to identify and verify common conditions based on patient presentation. This also suggests that clinicians do not necessarily realise the struggle that hypothesis generation represents for students. Two students in year 5 added that formulating a management plan is also a difficult aspect of the clinical process for them. Indeed, this is one of the key learning objectives in year 4 and 5 at the Southampton Faculty of Medicine. However, this aspect is beyond the scope of this research, which is mainly focused on history taking.

Students and clinicians generally agreed that virtual patients can be useful tools to foster diagnosis and decision-making skills. When making a diagnosis, students struggle to identify the conditions to test for and to adjust their diagnosis using new information. Clinicians attribute this to difficulties interpreting information gathered through the diagnostic assessment process, and evidence from the literature could explain these results using scheme theory, as described in Chapter 4.

Useful Types of Feedback

A central aim of this preliminary study is to determine the types of feedback that students find most useful in a self-learning and self-assessment virtual patient system. The usefulness of each feedback type has been evaluated on a 6-points Likert scale, ranging from "Completely useless" to "Very useful". The participants' responses for each feedback type have been analysed, comparing results between clinicians' and students' answers by year group. Due to the small number of participating clinicians, no significant comparative analysis could be performed between year groups for clinicians. Table 5.1 shows a comparison between students' and clinicians' ratings for each feedback type.

	Students			*Clinicians*		
	Negative	*Neutral*	*Positive*	*Negative*	*Neutral*	*Positive*
Feedback Type 1: Incorrect Actions	8(17%)	16(31%)	27(52%)	0	9(56%)	7(44%)
Feedback Type 2: Correct Actions	4(8%)	7(14%)	40(78%)	0	1(6%)	15(94%)
Feedback Type 3: Coherent Diagnosis	4(8%)	15(30%)	32(63%)	1(6%)	4(25%)	11(69%)
Feedback Type 4: Order of Actions	5(10%)	20(40%)	25(50)%	0	7(44%)	9(56%)
Feedback Type 5: Ideal (Expert) Process	5(10%)	14(27%)	32(63%)	0	4(25%)	12(75%)
Feedback Type 6: Consequences of Actions	4(8%)	16(31%)	31(61%)	0	5(31%)	11(69%)
Feedback Type 7: Plausible Diagnosis	6(12%)	6(12%)	39(76%)	0	6(38%)	10(62%)

Table 5.1: Percentages of students and clinicians considering each types of feedback "Somewhat useless" or "Somewhat useful" (neutral) or "Useful" or "Very Useful" (positive)

The graph in figure 5.3 summarises the results for students.

The type of feedback considered most useful by both students and clinicians is feedback type 2. This type of feedback was rated positively by 40 students (78%). Fifteen clinicians (94%) also rated this type of feedback positively. Feedback type 7 was considered useful or very useful by a majority of students (76%), but only 62% of clinicians had a

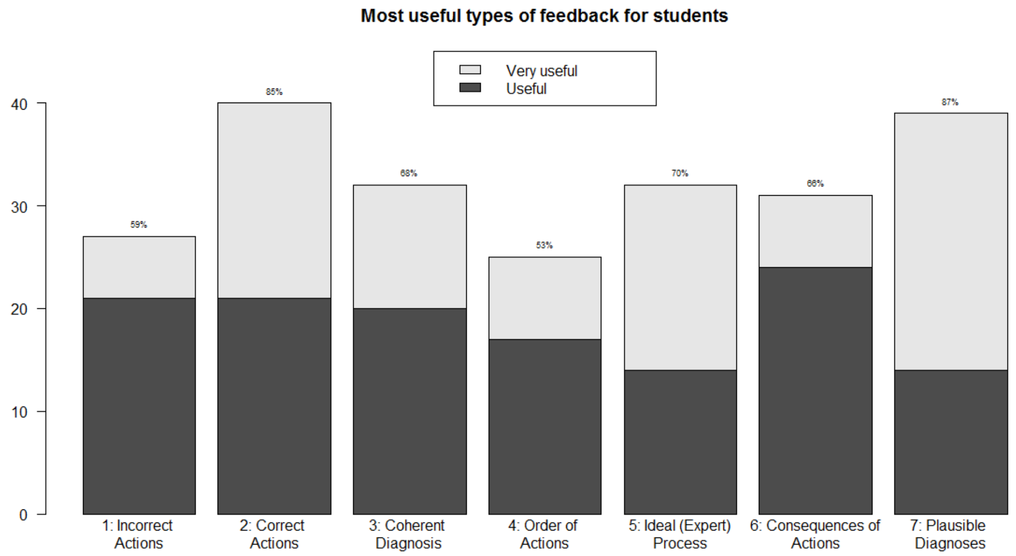


Figure 5.3: Types of feedback considered most useful by students.

similar opinion. This result underlines again the differences in reasoning between students with little clinical experience, and clinicians who encountered many cases throughout their practice of medicine, and are therefore able to formulate accurate hypotheses quickly using schemes elaborated through this clinical experience. As a result, clinicians may not be able to understand students' difficulties with hypothesis generation.

All other feedback types except type 4 were clearly considered useful or very useful by a majority of participant. Students and clinicians also suggested feedback regarding management options. This type of feedback could be seen as an extension of feedback type 6. Such feedback could contain narrative descriptions of what happens to the patient as a consequence of the student's treatment. Clinicians rated feedback type 3 and 6 most highly. These results could indicate that clinicians are concerned with students' thought process, while students seem more worried about generating appropriate diagnoses and knowing which actions are most appropriate. This again could be explained by scheme theory applied to clinical reasoning. Clinicians are more concerned with high-level reasoning, since their medical knowledge is appropriately encapsulated for clinical practice. Students, however, still need to reason on a lower level.

Opinions Regarding the Proposed Virtual Patient System Model

Three models of interactions have been proposed to both students and clinicians, and they were asked to select interaction models they considered useful for a self-learning and self-assessment virtual patient system. Participants were allowed to choose several models if desired. The resulting answers are detailed as follows:

- *Option 1:* "Being guided step by step throughout the case with questions and quizzes with feedback on each question (linear case)". This model has been selected by 36 students (70%) and 7 clinicians (45%). Students, especially in year 3, require close guidance when it comes to clinical skills, while clinicians feel that they need to be more independent.
- *Option 2:* "Having a limited number of choices and seeing the consequences of each choice until the case ends in success or failure (branching case)". This model has been selected by 25 students (49%) and 10 clinicians (44%). This shows mixed opinions about the branching model.
- *Option 3:* "Being able to make as many decisions as possible unguided and obtaining a global feedback on each choice made at the end of the case (exploratory case)". Only 20 students (40%) and 4 clinicians (25%) have selected this option. This result challenges assumptions concerning the best models of interactions for virtual patients. Having too many choices of actions within the virtual patient does not seem to be an important feature to most students or clinicians. This suggests that offering only a limited list of actions to perform in a virtual patient could be sufficient for the purpose of clinical self-study. However, to choose which actions students should be allowed to perform in a virtual case, virtual patient designers need to consider the pedagogical objectives of each case carefully.

A simplified model illustrated in figure 5.1 was shown to students and clinicians. 30 students (58%) and 12 clinicians (73%) consider that this model corresponds to the way they think about a clinical case. 29 students (56%) and 8 clinicians (50%) consider that the model contains all necessary information to make an accurate basic diagnosis.

A majority of students and clinicians favour an interaction model with a limited number of possible actions to choose from, and consider that the proposed simplified model of interaction corresponds to the way they think.

5.3 Paper Prototyping and Interviews with Students

Participants to the survey mostly agreed that that the high-level model of interaction proposed in figure 5.1 corresponded to their understanding of the clinical process. This section presents the interviews conducted in order to prepare the interaction design for SemVP (see Chapter 6) on a more detailed level. The primary objective of these interviews was to identify any major usability problem in SemVP's interaction design. To achieve this, paper models were used to represent SemVP's interface during the interviews. The secondary objective was to observe the students' thinking process while they were solving two given clinical case: decisions made, questions asked and examinations performed, etc.. The interviews were based on two example cases downloaded from the Electronic Virtual Patient project (eVip):

- Ms. Matibunda, a 67-year old lady who complains about an increased sensibility and painful fingers when she takes a tight grip on objects. The correct diagnosis is diabetes mellitus type 2. This case was developed by the Department of Educational Development and Research at the University of Maastricht.
- Catherine M., a five month old infant with fever and a reduced general condition. The correct diagnosis for this case is bacterial meningitis. It was developed by St George's, University of London.

Both cases are available under a Creative Commons licence. Four medical students were individually interviewed, among which 2 were in Year 3 and 2 in Year 4. Students were interviewed twice, for an hour each time, in a meeting room at the Southampton Faculty of Medicine.

First Interviews

During the first interviews, the strategies used by students to obtain a diagnosis were identified through protocol analysis, a structured interview technique used for knowledge engineering and ontology design (Shadbolt, 2005).
In the beginning of the first interview, the virtual case scenarios were presented to students, using only the short introduction text provided in the Medbiquitous XML files. For Ms. Matibunda's case, the presentation read as follows: "A 67-year old lady of Suriname background. Height 169 cm, 57 kilograms and a waist circumference of 73 cm. She complains about increased sensitivity". Catherine M's case was presented using the following introduction: "Ms. Miller comes to the outpatient department of the children's hospital with her five month old daughter, Catherine. Catherine has had a fever for two days and has become increasingly more weak and flaccid.".

Each case was discussed in turn, and explored using the following interview framework:

- *Possible diagnoses considered*: Can you think of any likely diagnosis given the information you have at this point? Why? Which factors in the patient's presentation make you consider these diagnoses?
- *Actions and decisions*: What would you do next (examination, interview question)? Why? What do you want to test by performing this task? What would you do depending on the outcome of the task? What sort of feedback would you need about this task?
- *Inferences from history taking and examination results*: Once you know the outcome of this task (the outcome of each task was provided using the information

contained in the existing Medbiquitous file), what would you do next? Does it change your initial diagnosis? Why?

These interviews were recorded using a digital audio recorder, transcribed and analysed after the interviews.

Second Interviews

During the second interview, students were presented with a paper mock-up of the virtual patient interface, and were invited to use it to demonstrate how they would interact with the virtual patient. To simulate how they would use SemVP's interface, students pointed at elements that they wanted to click on, and the researcher added new pieces of paper representing the changes that would appear on the interface as a consequence (see Snyder (2003) for details on paper prototyping techniques).

The workflow of students' activities in SemVP is described, screen by screen, on the UML activity diagram in figure 5.4. It is closely mapped to the workflow submitted to students and clinicians for evaluation in the survey (see figure 5.1).

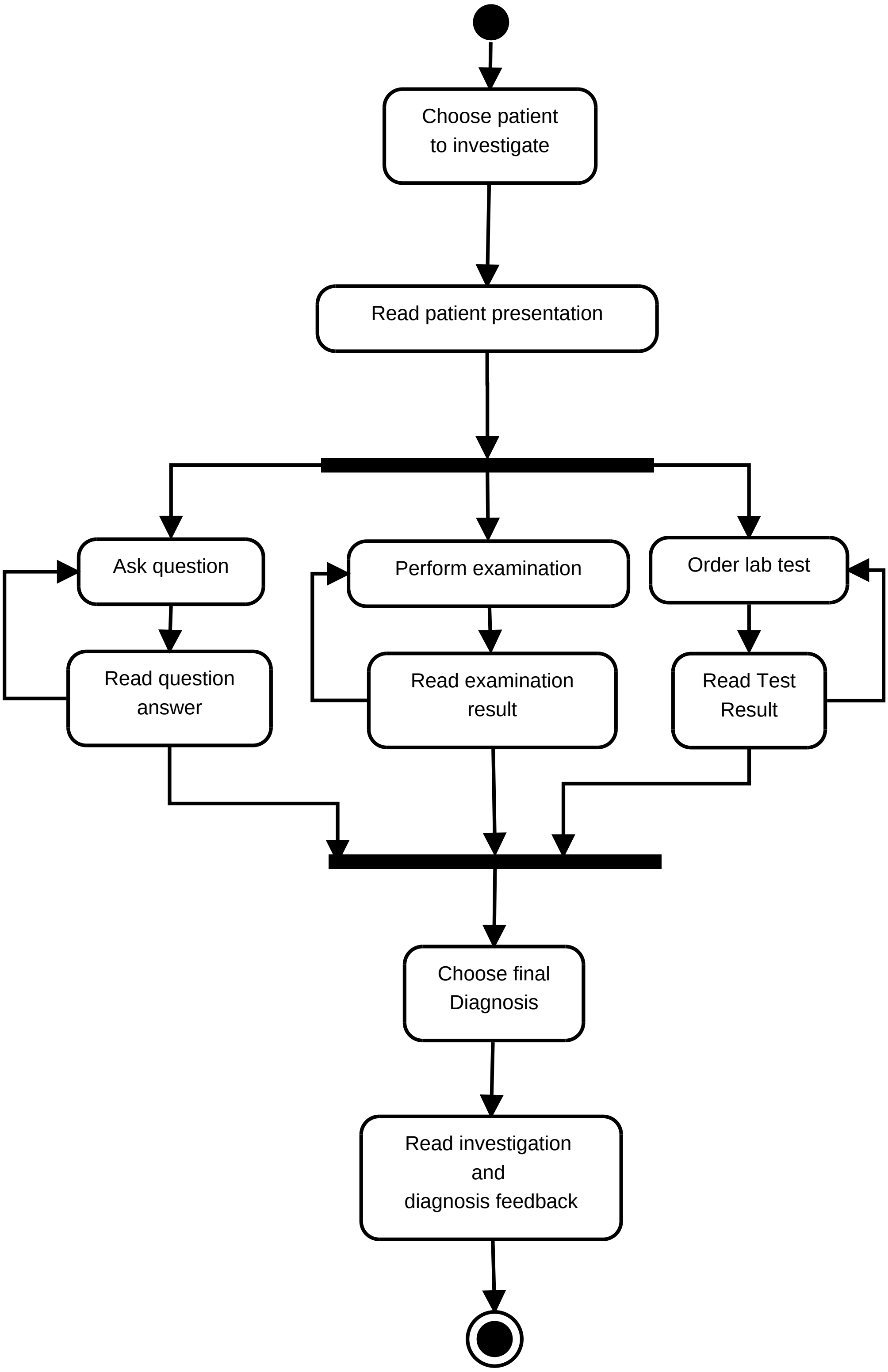


Figure 5.4: Virtual patient workflow

The process can be described as follows:

1. The student logs in to the application (the login page is not represented on figure 5.4 for simplicity), and chooses a patient to work on,

2. The student reads a brief description of the patient's main complaint or presentation, and navigates to either the *Questions* page, the *Examination* page, or the *Laboratory Tests* page,

3. On each page, the student selects the relevant action(s) they want to perform and obtain the corresponding answer or result,

4. Once the student has chosen a final diagnosis, the choice can be selected using the *Diagnosis and Management* page, and the feedback is then displayed.

As they manipulated the mock-ups, students were invited to think aloud and discuss their decisions using the following framework:

- *Possible diagnosis considered*: Can you think of any likely diagnosis given the information you have at this point? Why? Which factors make you consider these diagnoses? How would you expect to enter the diagnoses you are thinking about in the system?

- *Actions and decisions*: What would you do next (examination, interview question)? Why? How would you expect the system to allow you to perform this task? What outcome do you expect from this task? How would you expects the results of this task to appear on the system?

- *Inferences from history taking and examination results*: Once you know the outcome of this task, what would you do next? Does it change your initial diagnosis? Why? How would you expect to take notes of these reflections in the system?

- *Treatment and management*: Once you have reached a diagnosis, how do you expect to enter it into the system? What sort of feedback would you expect from the system once you have entered your proposed diagnosis?

These interviews were recorded using a camcorder on a tripod, pointed at the paper model itself and the students' hands manipulating it. This setup allowed an audio recording of the students' interviews, as well as an analysis of their actions on the paper prototype, in conjunction with their comments. Using the results of these interviews, SemVP was implemented for the final experiment presented in Chapters 7 and 8.

5.4 Interview Results: Students' Thought Process and Interface Usability

From the questions and examinations chosen by students, two main categories emerged: general history taking questions (generally used for hypothesis generation), and more focused questions aimed at ruling out or confirming certain conditions (hypothesis verification). General history taking questions start from the patient's presentation, and allow students to assess their general characteristics. A practical acronym to remember broad categories for these questions is S.O.C.R.A.T.E.S:

- *Site*: where, local/diffuse, "Show me where it is wors",
- *Onset*: rapid/gradual, patterns, "when did the symptoms begin?",
- *Character*: vertigo/lightheaded, pain: sharp/dull/stab/burn/cramp/crushing, etc.,
- *Radiation*: "does it hurt on both hands?", "does it hurt on all fingers?", etc.,
- *Alleviating factors*: "What makes it better?",
- *Time course*: when last felt well, chronic, etc.,
- *Exacerbating factors*: "What makes it worse?",
- *Severity*: on a scale of 1 to 10.

5.4.1 Case 1: Ms. Matibunda

The case was initially described as follows, using text provided in the original Medbiquitous XML file: "Ms. Matibunda, a 67-year old lady who complains about an increased sensibility and painful fingers when she takes a tight grip on objects.". The initial complaint from the patient is the following: "My fingers hurt when I take a strong grip on things, and it seems I feel everything on my skin more intensely. It worries me. It just does not feel right. And you hear a lot of things these days"
Students started with general history taking questions, and later moved on to more specific enquiries, as they started to suspect a given condition or type of conditions.
Students started by asking what the increased sensitivity involves, and then moved on to general history taking questions such as:

- "Where is the pain located? Can you feel it anywhere else in the body?"
- "How bad is the pain from (0 to 10)?"
- "When did the symptoms start?"

- "Does anything make the condition better or worse?"
- "Is it worse/better at any time of the day?"

After a few general questions, students quickly formulated hypotheses concerning the patient's condition, and asked more focused questions to confirm or rule out their hypotheses. Suspected conditions included: lung tumor, arthritis, neurological conditions such as multiple sclerosis (MS), and diabetes.
Relevant feedback regarding these hypothesis could include indicating the relevance of each of these hypothesis, using the associated symptoms and characteristics included in the patient's data (see Chapter 4).

- Concerns about lung tumor came from the idea that such a tumor could press on the nerves in the brachial plexus, thus creating pain in the upper limbs, bilaterally. Students asked questions to check if the pain was bilateral (on both hands) or unilateral. Despite the answer (bilateral pain), students did not investigate this possibility any further, and only asked if the patient was a smoker, which was not the case. After this answer, students did not investigate this hypothesis anymore, but focused on more likely diagnoses.
- Investigation concerning arthritis included looking and asking for any stiffness, swelling, and other signs visible on the hands.
- Neurological conditions were investigated in detail, and students enquired about the following aspects of the patient:
  - Blurred vision,
  - Unusual memory losses,
  - Changes in behaviour,
  - Family history of MS,
  - Muscle weakness or muscle atrophy.

  Students also suggested performing a cranial nerve examination and a neurological examination.
- Diabetes (the correct diagnosis) was also investigated by all students in detail, focusing on the following areas:
  - The three main symptoms of diabetes: weight loss, polyuria (excessive urination) and polydipsia (excessive thirst),
  - Family history of diabetes,
  - Vision problems,
  - Heart issues,

– Malfunctioning kidneys,
– Pain and numbness in legs and feet, typical for diabetes patients,
– Cholesterol (risk factor in diabetes),
– Hypertension,
– Check for glucose in blood and urine (blood test and urine dipstick).

Students also suggested performing a full blood count to check for any other anomalies.

Each hypothesis formulated by students led to a specific set of questions. Useful feedback could describe the relevance of each question in regards to the proposed hypotheses. For instance, if diabetes is suspected, questions regarding family history are particularly relevant.

After conducting all these investigations, no interviewed student was able to determine the correct diagnosis. Indeed, the data used in the original XML file was not representative of a typical diabetes patient, but described a highly unusual presentation of the condition. These results might have been misleading for students. However, the purpose of these interviews was not primarily to determine if students would find the correct diagnosis, but to investigate the choices they made based on the patient's initial condition. Students did indeed investigate the possibility for diabetes in depth, as well as other conditions. They also investigated other likely conditions in detail, according to the estimated probability of each one. When students ruled out likely diagnoses such as diabetes and neurological conditions, they returned to more general questions such as previous hospital admissions, previous surgeries, general health history and daily habits (smoking, diet, etc.), until they ran out of options and the interview stopped.

These interviews highlight a process of general history taking, focusing first on a broad assessment of the patient's main symptoms and complaints, followed by more focused investigations as hypotheses arose. Once the initial diagnosis propositions were eliminated, students returned to a more general inquiry, looking into the patient's medical history, environment and general wellbeing.

5.4.2 Case 2: Catherine M.

Students were presented with Catherine's case as follows: "Ms. Miller comes to the outpatient department of the children's hospital with her five month old daughter, Catherine. Catherine has had a fever for two days and has become increasingly more weak and flaccid.". The setting of this case differs from the previous one in that the situation occurs in a hospital, instead of a GP surgery. Moreover, the situation is more worrying, potentially fatal, and rapid patient management is a priority.

When asked about their hypotheses in terms of diagnosis and the next course of action they would choose, students immediately cited meningitis as a probable diagnosis. Subsequently, the chosen interview questions and examination were initially highly focused on this specific diagnosis. Indeed, in this situation the priority would be to check the patient's vital signs, and to confirm and treat meningitis as soon as possible if appropriate, to avoid potentially fatal complications. The students' responses reflected this priority.

Relevant feedback in this case would include listing critical actions missed by students. These actions, if not performed, could lead to critical consequences for the patient (feedback type #6 could be used to show this, see Section 4.3).

Initial interview questions asked by students included:

- "Is Catherine drinking normally? Has she been drinking less lately?" (student also stated that they would check for clinical signs of dehydration at that stage, instead of relying on Catherine mother's assessment. Catherine shows definite signs of dehydration).
- "Is Catherine coughing?" (there was no coughing in this case).
- "Any rashes or other abnormal patches on the skin?" (no such symptom was present in the case).
- "Is Catherine holding herself in an unusual position (arching, struggling to hold her head straight)?" (Catherine does hold herself in such a position, but students moved on to the next questions without waiting for an answer to this question).

At that stage, students considered meningitis less likely, in particular bacterial meningitis, because of the long time since the symptom started (2 days, as indicated in the case's introduction). This means that students changed their opinion concerning the case without using the answers to the above questions. In this situation, feedback type #2 would be particularly relevant, reminding students of the importance of using all the information they gather.

Students then proceeded to more general questions, in order to confirm or rule out the various possible types of infections such as urinary tract infection or otitis. Meningitis was still considered a possibility in this investigation, given the severity of the prognosis for such a condition.

Students inquired about the following topics:

- Runny nose,
- Abnormal quantity and aspect of stools and urine,
- Trauma, such as downfall or impact on the head,

- Similar symptoms in the family,
- Crying,
- Vaccinations,
- Vomiting,
- Discharge of unusual substances from the ears,
- Pregnancy, birth and growth of the child,
- Preexisting conditions, previous admissions to the hospital,
- Allergies.

Subsequence actions included checking for infections using laboratory tests and examinations such as:

- Checking for abnormal breathing sounds,
- Blood, stools and urine cultures,
- Sceptic screen, including blood test, chest X-ray, lumbar puncture and midstream urine analysis.

These interviews, set in a context were fast patient management is important, highlighted a different way of assessing the patient's health, starting with the assessment of a single and potentially life-threatening condition. This condition is investigated in priority, using all the corresponding means of investigation, without dwelling on the more general aspects of the patient's history initially.
The patient's general health history is investigated later, when students mostly ruled out the initial diagnosis.

5.4.3 Interaction and Usability

Simple paper models representing a proposed virtual patient interface were designed and presented to the interviewed students. These models were used to simulate how students would use the virtual patient system, and thus identify any major usability problem before implementation.
Students were invited to show how they would use the interface model to perform the following tasks:

- Asking a question,
- Examining the patient's hand,

- Adding a diagnosis to a list of possible hypotheses,
- Submitting a final diagnosis.

The interviewer modified the paper model by adding new sheets of paper or writing on the prototype according to students' actions, mimicking the behaviour desired from the proposed system. Students were also encouraged to "think out loud" and to describe any hesitation they might have, any element of the interface they might not understand, and more generally anything that might prevent them for achieving the task efficiently.

Asking a Question To The Patient To test how students would ask a question to the patient, they were presented with paper model showed on figure 5.5.

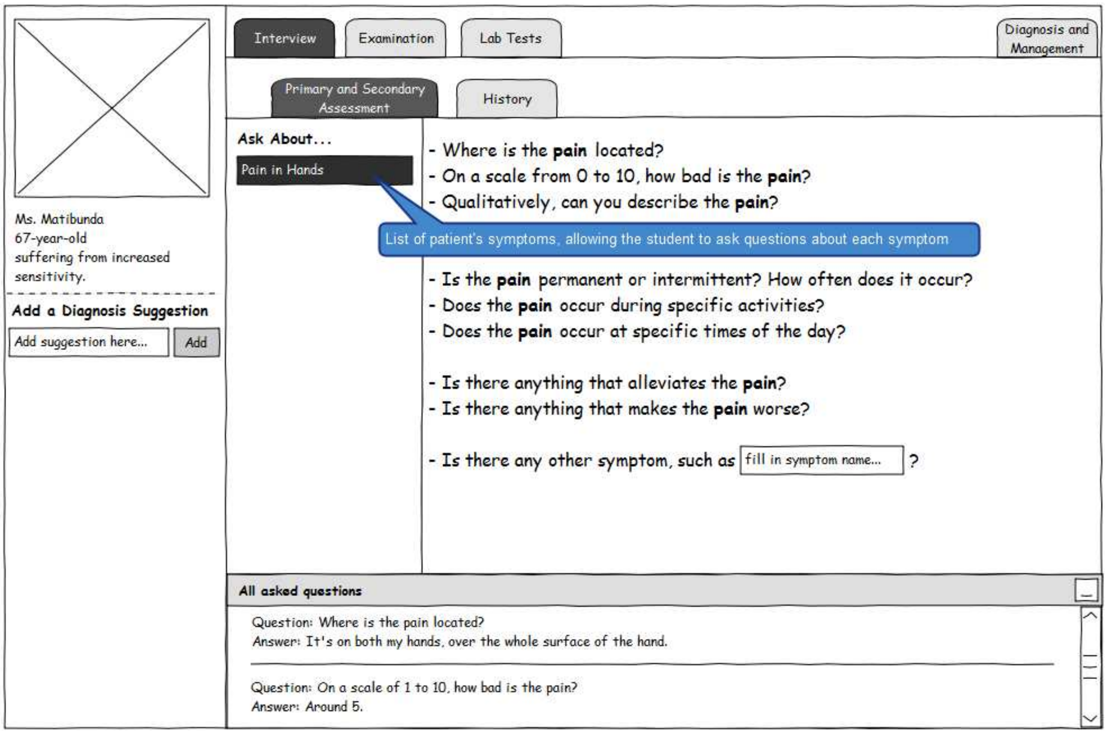


Figure 5.5: Paper model: "Interview" page

This screen features several lists of questions grouped by theme. The left columns of the screen features a picture of the patient and a very brief description of the case, as well as a section allowing students to suggest new diagnoses at any point in the case. Based on the categories of questions identified in previous interviews, questions were grouped in two sub-tabs:

- *Primary and Secondary Assessment*, containing all questions directly related to the patient's complaints and symptoms.

- *History*, containing all questions related to the patient's general health.

Clicking on each symptom reveals questions concerning each symptom's characteristics, such as duration, radiation, severity, etc.. In the second sub-tab, several general health categories are laid out: Developmental Milestones, Occupation and Environment, Medical History, etc.. Students could click on each category to reveal the questions related to the category. The lower part of the screen is a list of all questions previously asked by students, with all corresponding answers. Students agreed that this feature was a useful reminder of all previously asked questions.

Even though students could perform the required task (asking the patient about the location of the pain), they noted that using these categories with such a complex navigation system was too complex. The general consensus was that a much simpler interface would be more appropriate. For instance, students expected to type "hand pain" in a search form, and obtain a list of all questions related to the patients pain in the hand.

Examining the Patient The next task required from students was to examine the patient's hands. Doing so required using the "Navigation" tab on the top of the page to access the examination page (figure 5.6). The examination page features a top section with a list of general measurements such as heigh, weight, temperature, etc. On the lower section of the screen, a picture of the patient is displayed, enabling students to move a colored square to the anatomical area to be inspected, then selecting the action to be performed for the examination (Inspect, Percuss, Listen, Functional Test, Measure and Palpate).

Figure 5.6: Paper model: "Examinations" page

Students were also able to understand and accomplish the task, and in this case too they suggested a simpler interface.

Adding a diagnosis to a list of possible hypotheses On the left panel of the interface, a simple form was proposed, enabling students to add diagnoses that they suspect for the patient. Students could add suspected diagnoses using an "auto-fill" form, as demonstrated on figure 5.7. When a diagnosis is submitted, students could change their opinion regarding each proposed condition, using three radio buttons: "Ruled out", "Likely", and "Very Likely". Every change of opinion is recorded in the student's data, and used to generated feedback (see Sections 6.4 and 6.5). Indeed, students' assessment of the patient should evolve appropriately using the results of examinations and interview questions, and the evolution of the student's diagnosis can provide meaningful data for feedback generation.

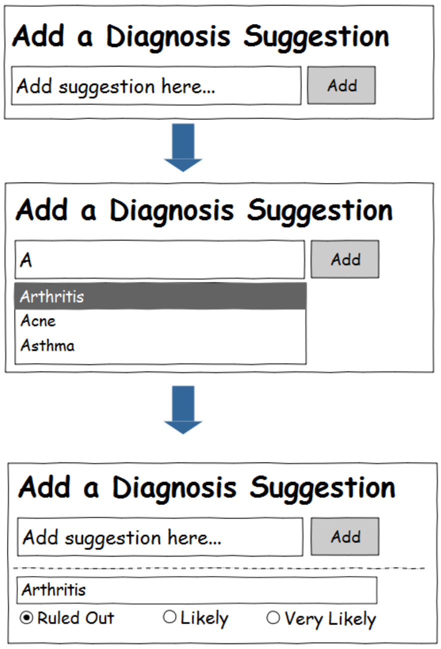


Figure 5.7: Suggesting a diagnosis: initial design (paper model)

Students understood this feature well, and had no difficulty to perform the task required ("You suspect Arthritis for this patient. Enter this information in the system"). One of the students even suggested including this feature to the system before being asked to perform this task.

Students were able to understand how to submit a final diagnosis using the simple list provided in the paper model.

5.5 Summary and Implications for the Design of SemVP

The results of the preliminary survey show that both clinicians and students see the pedagogical value of virtual patients. They also consider the proposed types of feedback to be generally useful, in particular feedback type 2 (appropriate actions to choose) and feedback type 7 (differential diagnosis). The most notable difference between students and clinicians is in opinions regarding the difficulties students experience and the corresponding feedback needed to help students improve. Students struggle most with hypothesis generation, which can be explained by script theory (see Chapter 4). Since students lack extensive clinical experience, they are unable to use their biomedical knowledge in a clinical context, and struggle to generate targeted hypotheses. Experienced clinicians, on the other hand, are able to use schemes elaborated through many encounters with patients, combined with preexisting biomedical knowledge, to generate targeted hypotheses quickly. As a result, they do not always understand why students struggle with hypothesis generation.

From the interviews, situations emerged were the relevance of feedback type #2 and #7 would have been relevant to help students understand their mistakes and evaluate their choices.

Benefits of the semantic web for feedback types 2 and 7

Feedback types #2, in current systems, can only be delivered as a free text list of relevant diagnosis tests to perform. However, this type of feedback is generic and disconnected from each student's path in the clinical scenario. Providing personal feedback directly related to what the student has done or has failed to do would promote a deeper reflection for students. A semantic model detailing each student's action could generate a personalised list of relevant diagnoses that the student failed to consider, which would help each student to consider their own mistakes in a targeted manner.

Feedback type 7 could also be delivered in existing systems using free text, but using semantic web technologies allows the display of richer information, since conditions can be linked to students' choices and the patients' symptoms. This allows inferences about the diagnosis tests chosen by the student in relation to potentially relevant conditions, such as knowing if the student missed a relevant diagnosis test because she did not consider one of the plausible conditions.

As a result, a virtual patient system based on semantic web data will generate feedback types that most students consider useful, but will also provide richer information about each student's actions (feedback as "reaction"), which will enable students to reflect on their performance and enhance their reasoning skills.

Chapter 6

SemVP: a Semantic Web-Based Virtual Patient System

This chapter describes the design of SemVP, a virtual patient system prototype based on semantic web technologies. SemVP was designed to demonstrate how the Semantic Web can be used to generate automatic and personalised feedback. SemVP was evaluated during an experimental study described in chapter 7.
The semantic data used in SemVP incorporated OpenGalen[1], an open source biology and healthcare ontology written in OWL DL, providing numerous classes and properties that can be used to represent a virtual patient (Rector et al., 2003). Ontology design patterns, resources from knowledge bases such as Freebase[2], and *ad hoc* classes and properties have also been integrated to SemVP's semantic model, using the interoperability afforded by semantic web technologies. The resulting system supports the representation of the virtual patient and generates automatic feedback for each student.

6.1 Interface Design

The semantic model underlying SemVP was designed to be independent from any particular interface system. However, to evaluate the validity of the feedback generated by SemVP, a usable and clear interface was required. SemVP's interface was designed in HTML and CSS. The design was based on the interviews conducted during the preliminary interview described in Chapter 5.
The interaction model presented in the preliminary study(see figure 5.1) was used for SemVP's navigation structure. This structure enables access to various screens, which allows students to ask questions, examine the patient and order lab test (investigation).

[1] `http://www.opengalen.org/`
[2] `http://www.freebase.com/`

SemVp's "Interview" screen features a simple list of all possible questions on the left side of the screen (see figure 6.1). The questions were extracted from the original Medbiquitous XML files. Additional questions were included based on the interviews conducted during the preliminary study (see Section 5.4.3).

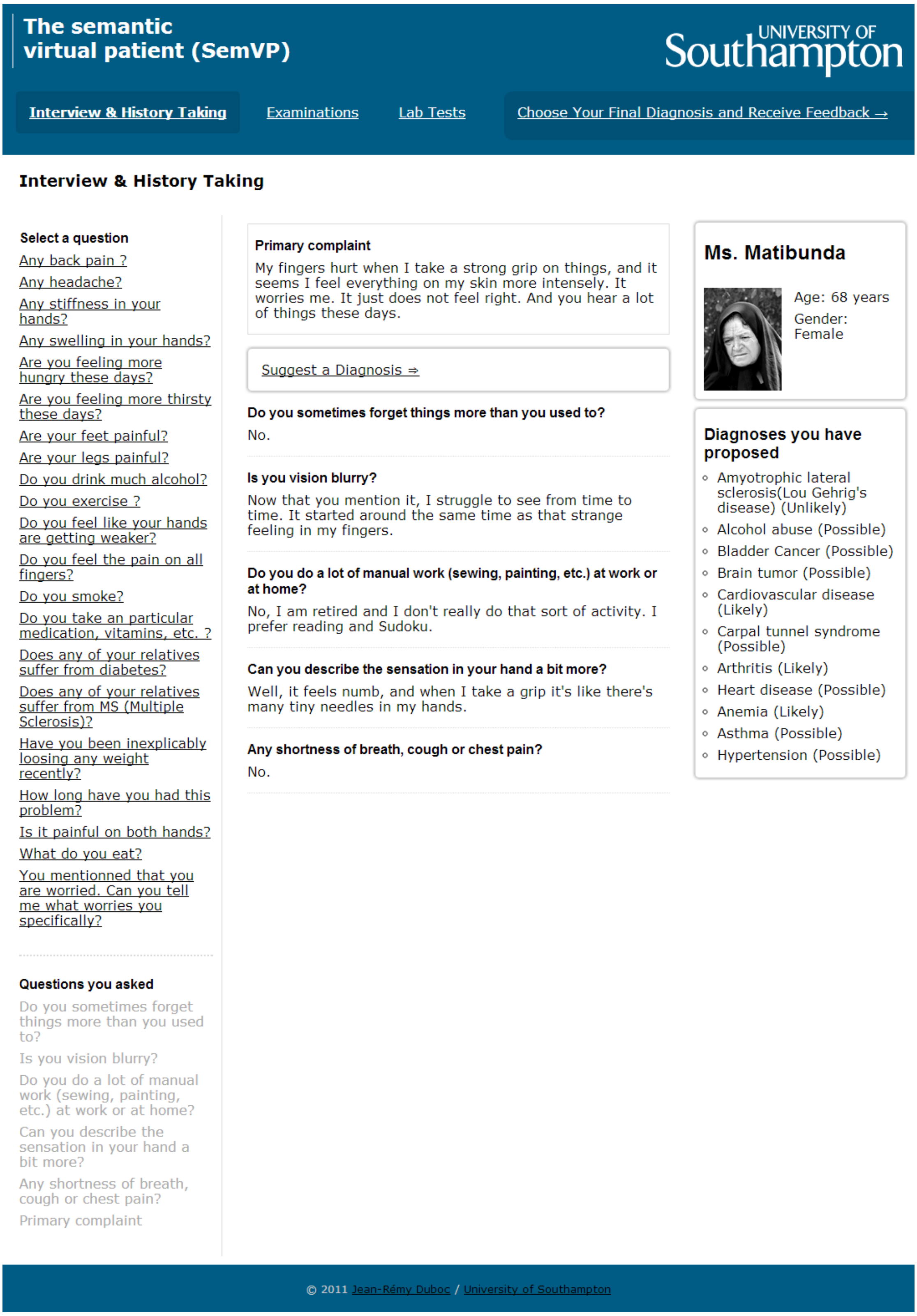


Figure 6.1: Final interface: "Interview" screen

SemVP's "Examination" screen features a list of available examinations on the left side of the screen, similar to the solution implemented for questions (see figure 6.2).

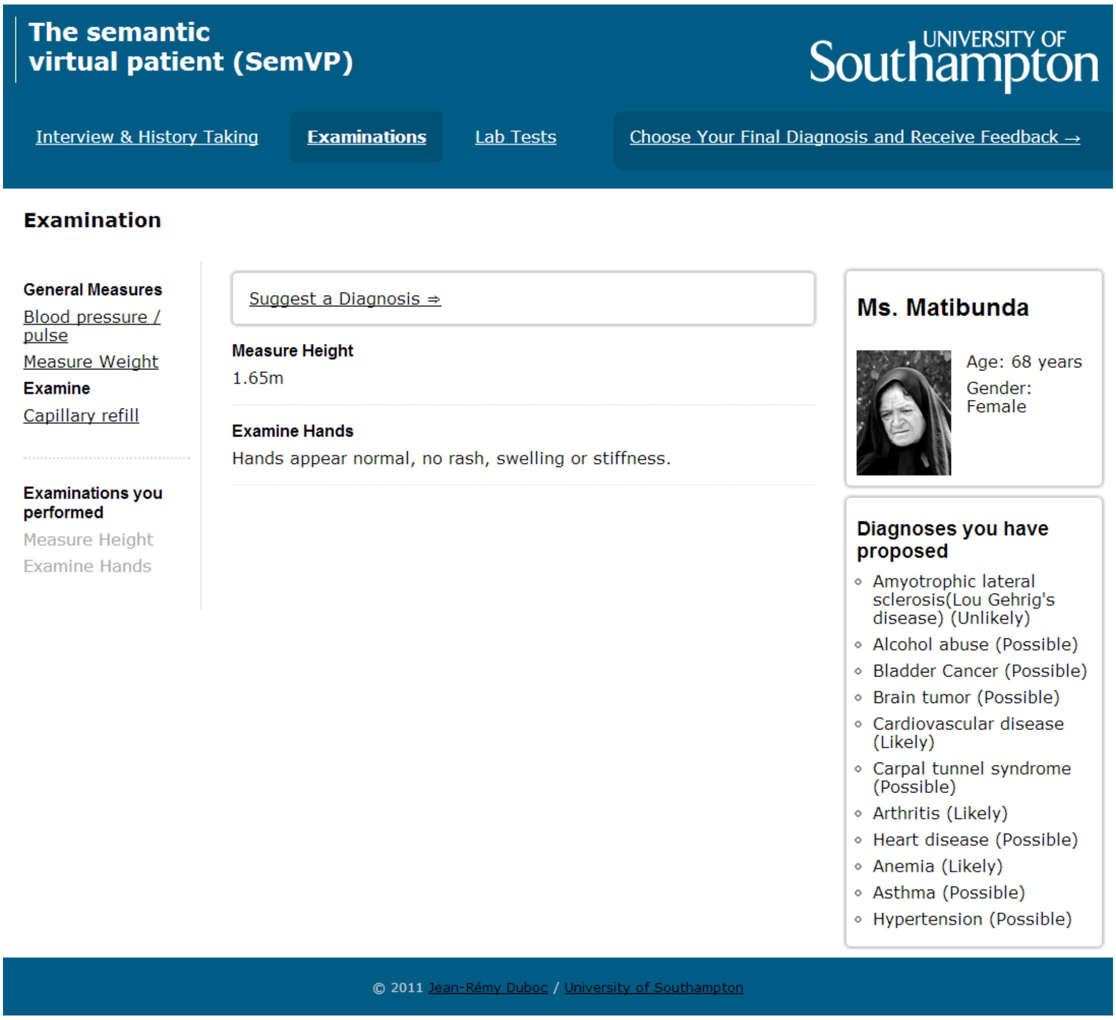


Figure 6.2: Final interface: “Examinations” screen

A similar solution was used for lab tests (see figure 6.3).

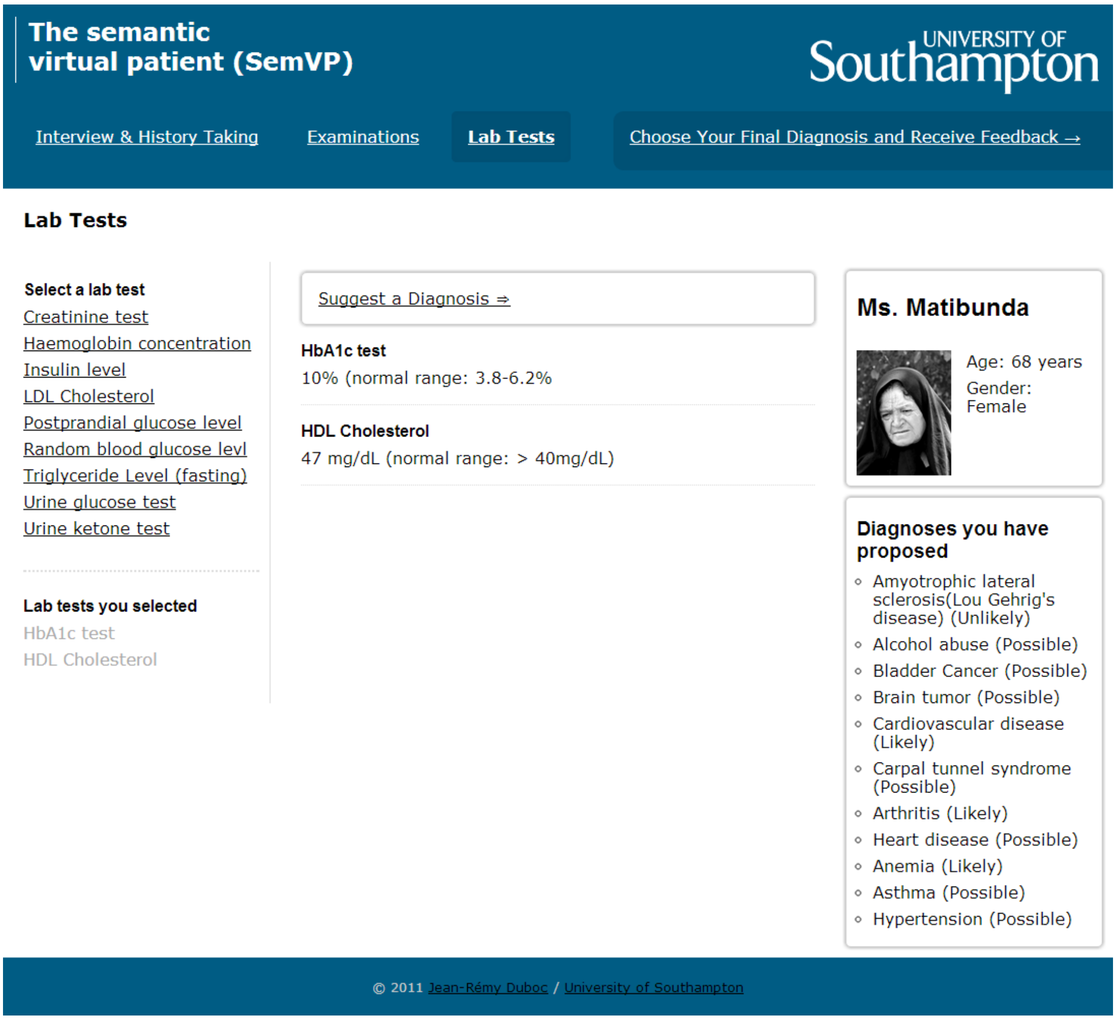


Figure 6.3: Final interface: "Lab Tests" screen

As described in figure 5.7, the initial paper model proposed to students used radio buttons to change the estimated likelihood for each suspected diagnosis. However, this model takes two stages for each proposed hypothesis: first, enter the suspected diagnosis, and then select the suspected likelihood for this diagnosis.
To simplify this process, SemVP's interface features a list of possible diagnoses that students can sort in three columns using a "drag-and-drop" motion (see figure 6.4).
As a result, selecting a possible diagnosis is done in a single step.

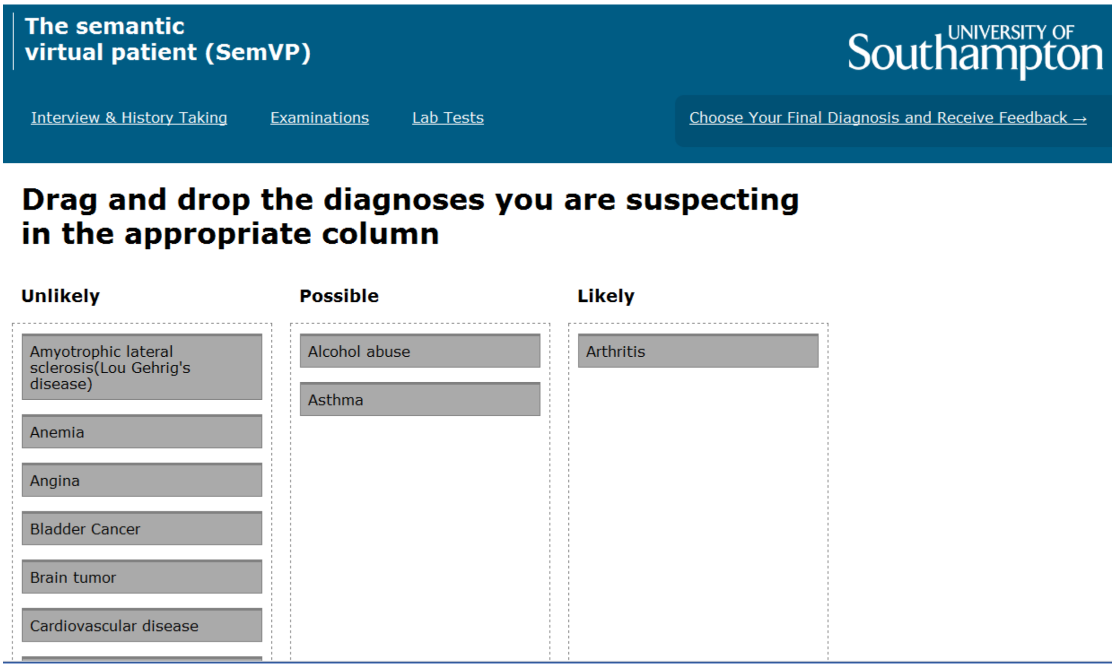


Figure 6.4: Final interface: diagnosis. Students can drag diagnoses from one column to another.

The interface displaying the final feedback is detailed in Section 6.5.

6.2 Technical Design

The SemVP system was implemented using the ASP.NET MVC (Model-View-Controller) framework[3]. The Model-View-Controller design pattern preserves the separation between SemVP's data model and its interface (Leff and Rayfield, 2001).
SemVP was deployed on a Windows server, with a Windows SQL server for the database. The SQL database was used to store user data, such as login and password, using the framework's default user management system. The high level structure of SemVP is represented in figure 6.5.

[3] `http://www.asp.net/mvc`

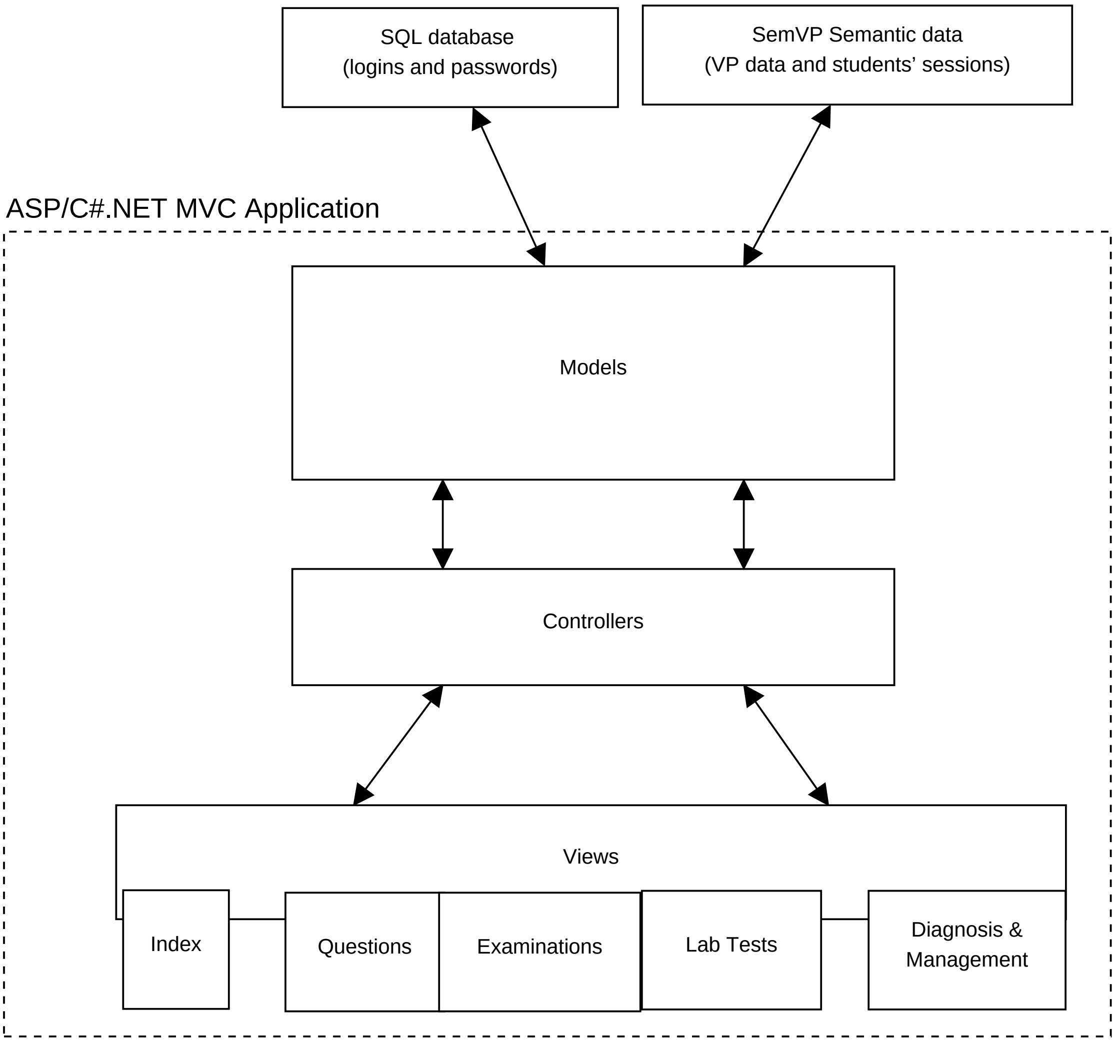


Figure 6.5: Model classes used in SemVP

The UML diagram in figure 6.6 shows how model classes are structured in SemVP's design.

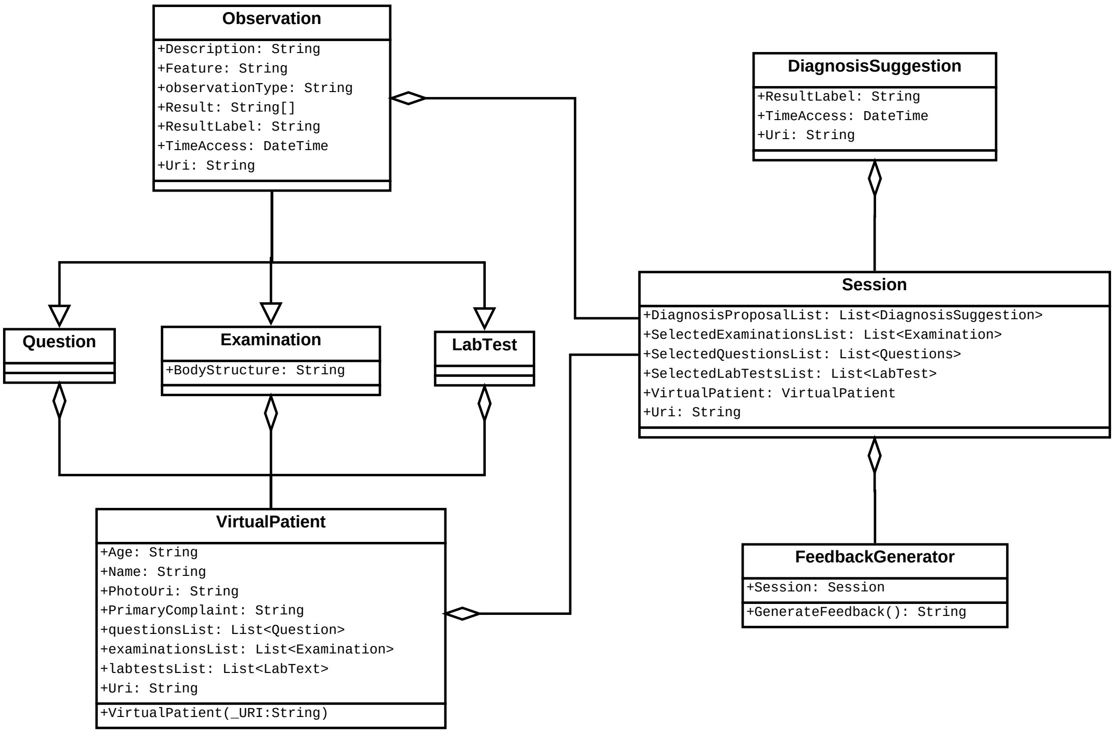


Figure 6.6: Model classes used in SemVP

Models are used to manage the semantic data representing the patient and the student's session. To do so, the following model classes were designed:

- The *VirtualPatient* class is used to retrieve data describing the virtual patient from the semantic data.
- The *Session* class is used to retrieve data describing the student's session, including all actions selected by students. A class called *DiagnosisSuggestion* is used to store and retrieved all suggestion made by each student.
- The super-class *Observation* is designed to retrieve all actions available to the student in the semantic data store. Three sub-classes are designed to retrieve the three types of actions available: *Question*, *Examination* and *LabTest*,
- The *FeedbackGenerator* class was designed to generate dynamic feedback for each student.

Each class retrieves data from RDF files using the dotNetRDF library[4].

Corresponding views and controllers were designed to allow students to interact with the model represented in the model classes. Figure 6.7 is a UML diagram representing SemVP's view and controller classes.

[4] `http://www.dotnetrdf.org/`

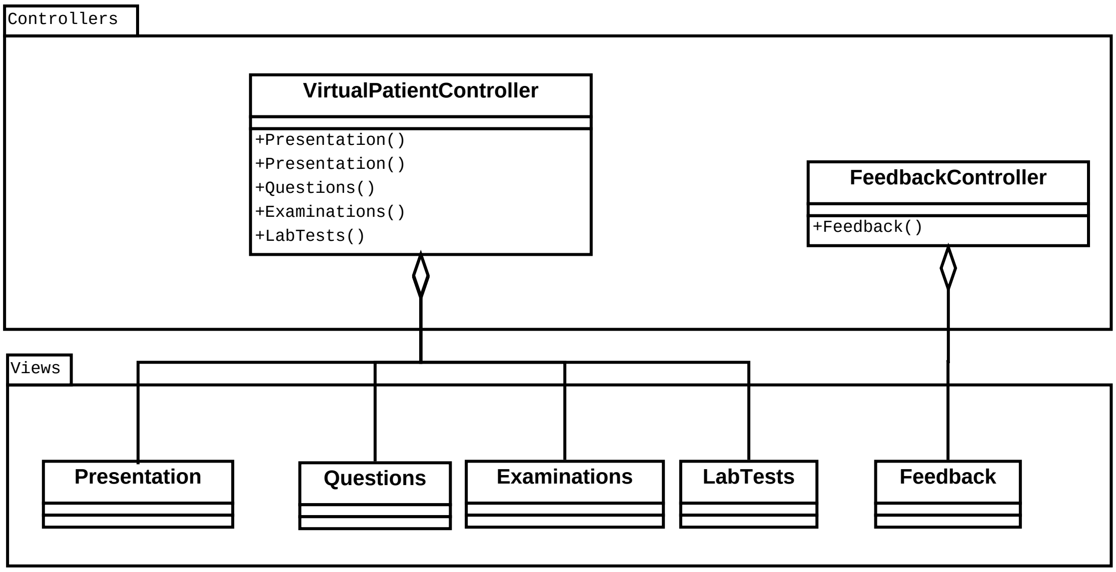


Figure 6.7: View and Controller classes used in SemVP

The first view class, called *Presentation*, displays the patient's information, such as age, gender, picture, and primary complaint. The three view classes called *Questions*, *Examinations* and *LabTests* handle the display of each question answer, examination result and lab test result. Finally, the view class *Feedback* enables each student to choose a final diagnosis and displays the feedback delivered by SemVP, described in Sections 6.5 and 6.6.

6.3 Representing a Virtual Patient in SemVP

To generate relevant feedback for students about their interactions with the virtual patient, SemVP uses a formal representation of virtual patients, based on semantic data. A virtual patient is represented by a node belonging to a class called *virtual_cases:VirtualPatient*. This class was created especially for SemVP. OpenGalen contains many useful classes that can be used to describe some aspects of the virtual patient more precisely. For instance, the node representing Ms. Matibunda (case described in Section 5.3) belongs to the class *opengalen:FemalePatient* class. SemVP also leverages classes from other data source, such as Freebase. Thus, the node representing Ms. Matibunda also belongs to the *dbpedia:Surinamer* class, which is part of the Freebase knowledge base. Defining the patient's gender and ethnicity in this way is especially helpful when these attributes are considered as risk factors for certain medical conditions. For instance, data from Freebase indicates that female patients are more susceptible to conditions such as thyroid cancer (see `http://www.freebase.com/view/en/thyroid_cancer` and `http://www.freebase.com/view/en/female`).

Chapter 3 describes how RDF can be used to connect resources (also called "nodes") from disparate online sources. This allows the integration of information from Freebase (such as the information described above) into SemVP's model. An example of how Ms. Matibunda's gender and ethnicity can be represented using data from Freebase is shown in figure 6.8.

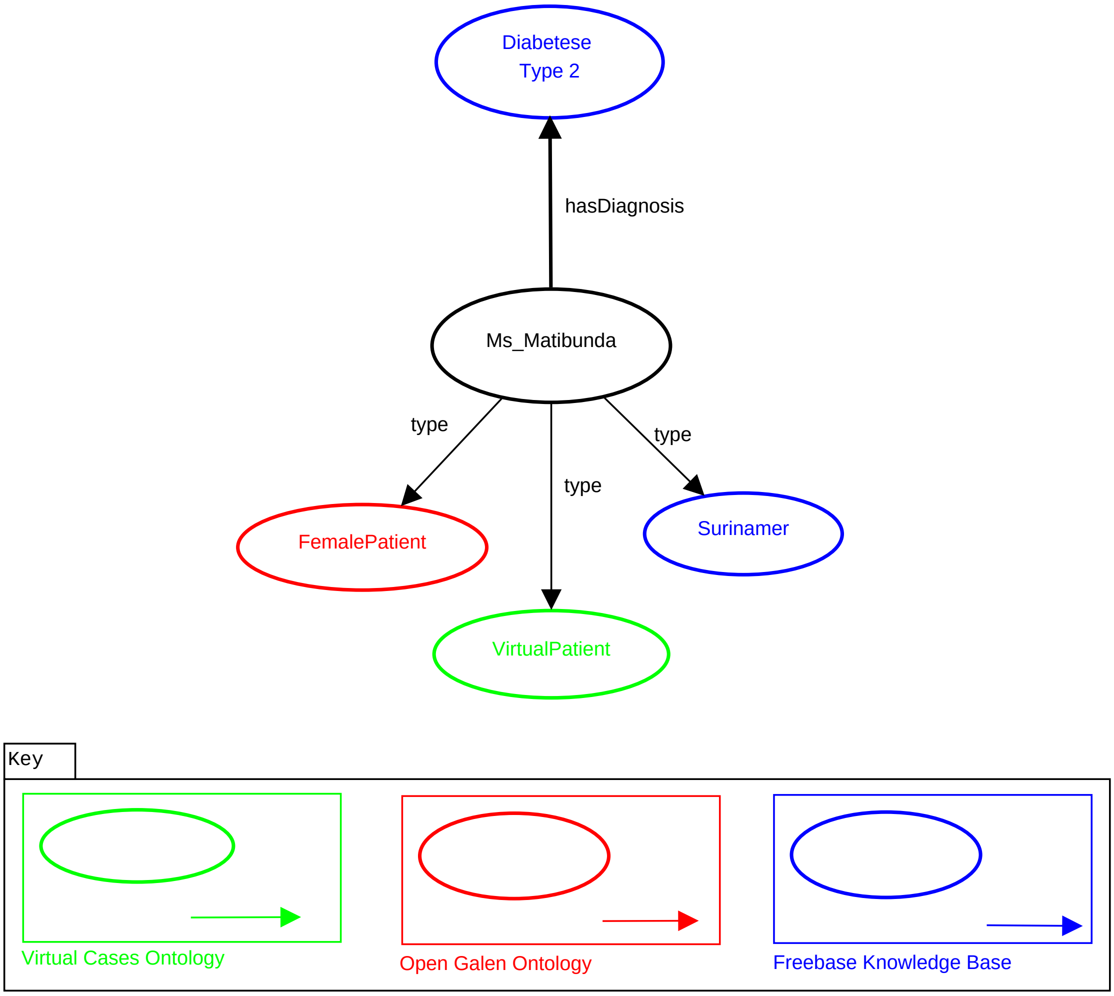


Figure 6.8: RDF representation of Ms Matibunda and the correct diagnosis

The RDF graph depicted in this figure shows that Ms Matibunda belongs to a class called *VirtualPatient* (created for the purpose of this research). Ms Matibunda also belongs to a class called *FemalePatient* (from OpenGalen), and to another class called *Surinamer* (from Freebase). Figure 6.8 also shows how the correct diagnosis (or diagnoses) affecting the patient is represented using the *hasDiagnosis* property from OpenGalen.

6.3.1 The Virtual Patient's Medical Features

The RDF graph in figure 6.9 shows how Ms. Matibunda's weight, age, and height are represented in SemVP.

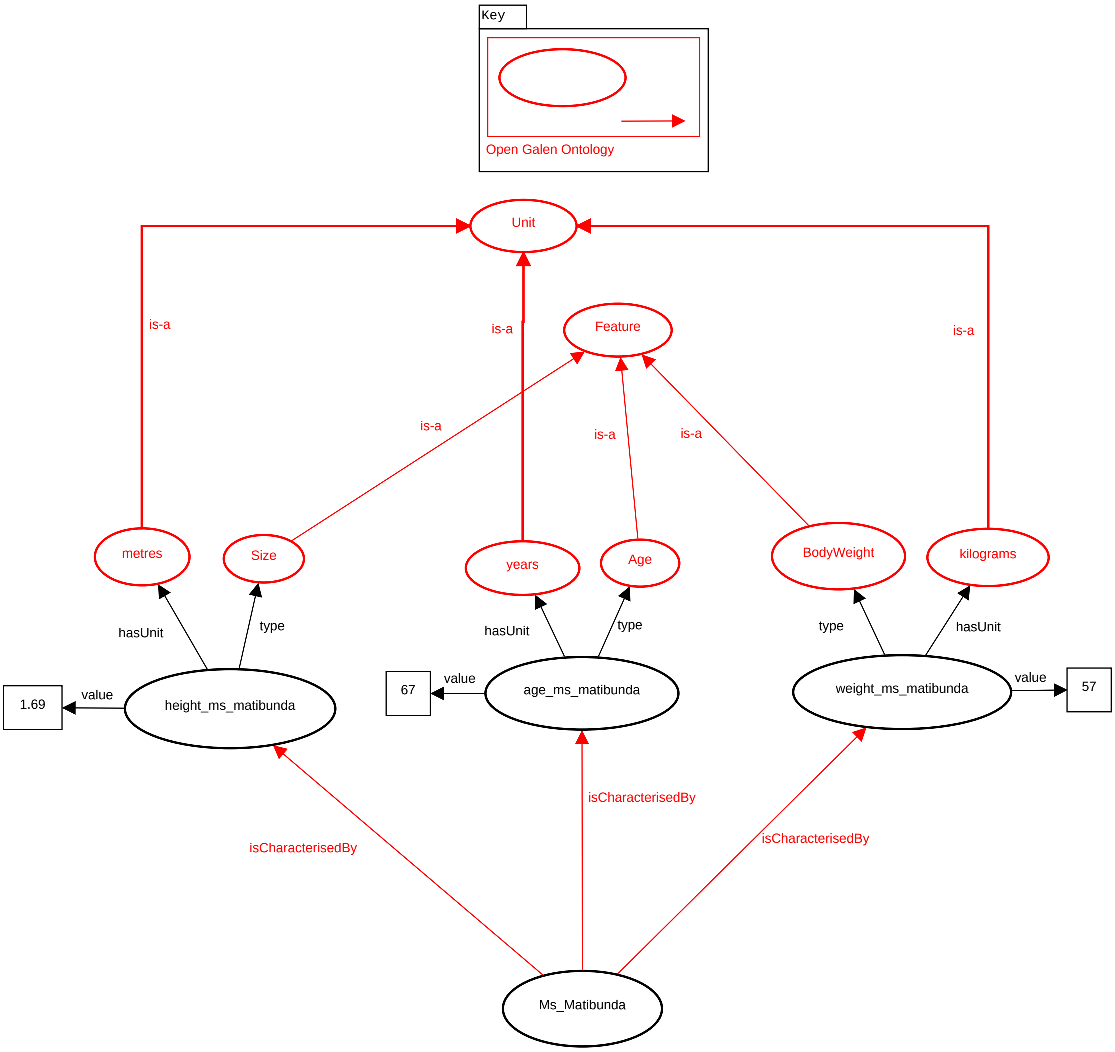


Figure 6.9: RDF representation of Ms Matibunda's general features

To represent a patient's physical characteristics (such as weight, age, height, etc.), OpenGalen provides a class called *Feature*. The *Feature* class refers to things that characterise the patient as a whole, or things that characterise a patient's body part. *Feature* has many sub-classes such as *StucturalFeature*, *OrganismFeature* or *Substancefeature*, each containing several sub-classes. Each of these classes allows the description of a specific feature. OpenGalen also provides two inverse properties called *isCharacterisedBy* and *characterises*, which enable a semantic link between the resource representing the virtual patient and the resources representing each one of its characteristic features.

It is also necessary to indicate the units used for each feature when needed. Numerous unit classes are available in OpenGalen and can be used for this purpose. For instance, to represent Ms. Matibunda's weight, the following statements are included in the knowledge base, as shown on figure 6.9:

- "*Ms_Matibunda* is characterised by *weight_ms_matibunda*",

- "*weight_ms_matibunda* has a value of *57*" (57 is an integer variable),
- "*weight_ms_matibunda* is of type *BodyWeight*" (*BodyWeight* is a sub-class of *Feature*),
- "*weight_ms_matibunda* "has unit" *kilograms*".

The use of RDF enables the representation of a large number of medical features by leveraging OpenGalen directly. These features range from name, categories, gender etc., to characteristics that evolve rapidly in time such as weight, height(size) and age.

6.3.2 Anatomical Description of the Virtual Patient

The OpenGalen ontology contains an *Anatomy* component, which contains classes designed to describe many human body parts. This component contains a class called *BodyStructure*, which contains many relevant subclasses, such as *BodyAsAWhole*. This class characterises the whole body, and therefore allows a description of general features concerning the patient. A large number of other body parts and anatomical entities are listed in the ontology. This allows the modeling of many individual body parts.
Body parts described in OpenGalen include, among others:

- Head and neck,
- Trunk body parts, such as abdomen, back, chest, etc.
- Extremities, including all fingers, listed under the class *Finger*, with subclasses *IndexFinger*, *LittleFinger*, *MiddleFinger*, *RingFinger*, and *Thumb*.

Two additional classes were included in SemVP, in order to distinguish between body parts situated on the left side and those on the right side of the body: *LeftBodyPart* and *RightBodyPart*.
Each body part can be characterised by a feature (described above), using the corresponding class *Feature*. This allows, for instance, the description of the size of any given body part.

6.3.3 Modeling Symptoms, Risk Factors and Conditions, and Linking Them Together

To provide useful feedback to students, SemVP uses two critical features:

- A list of conditions and symptoms, extracted from existing ontologies and knowledge bases on the web (Freebase for this thesis, but many other RDF knowledge base can be used),

- Links between conditions and symptoms, and links between conditions and risk factors.

This section describes how the design of SemVP caters for these features, using the OpenGalen ontology along with the Freebase knowledge base.

6.3.3.1 Describing Symptoms Affecting the Virtual Patient

The symptoms affecting the virtual patients constitute the most fundamental aspect of SemVP's model. It is the main thing that students need to investigate.
OpenGalen contains a class called *Symptom*. All nodes representing symptoms belong to this class. However, OpenGalen does not provide specific instances of symptoms. In this situation, the semantic web's inherent flexibility is a precious asset. As described in Chapter 3, RDF makes it possible to use symptoms available from Freebase[5] and to integrate them to SemVP's model.

An example of this is shown in figure 6.10 : the patient Catherine M. is affected by a headache and a fever. The symptom "headache" is pulled from the Freebase knowledge base. The URI for the headache is `http://rdf.freebase.com/ns/en.headache`. In this example, Catherine's temperature is both a feature representing Catherine and a symptom affecting her.
Any of the patient's medical features can be identified as a symptom or a risk factor of a given condition in this manner.

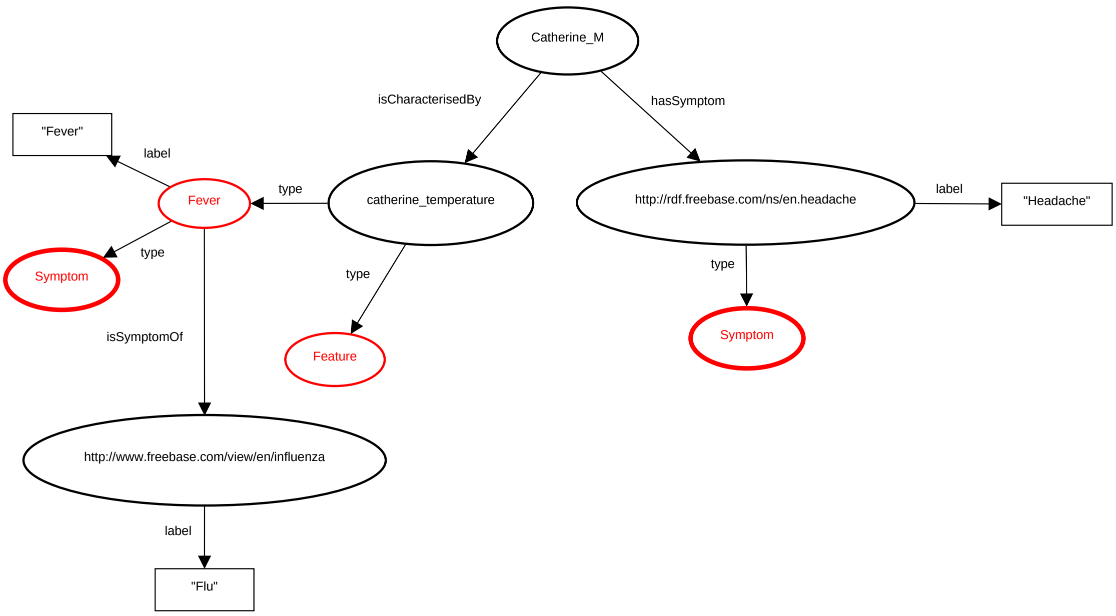


Figure 6.10: Example graph: link between a symptom pulled from Freebase and a virtual patient

[5] `http://www.freebase.com/`

It is also important to cater for use cases in which the patient is not affected by a given symptom. For instance, if the student asks Ms. Matibunda if she suffers from a headache, and Ms. Matibunda does not, the *absence* of headache needs to be formally described in the patient's data. For this purpose, a property called *isNotCharacterisedBy* was created.

6.3.3.2 Linking Conditions to Symptoms

Figure 6.10 features an RDF triple that reads as follows: "*Fever* is (a) symptom of *http://www.freebase.com/view/en/influenza* (flu)" (Chapter 3 describes what an RDF triple is). The property "*isSymptomOf*" is part of OpenGalen, and denotes a causal link between a condition and its symptom(s). However, OpenGalen does not provide individual instances of this property being used. As a result, the triple described above has to be authored manually depending on the specific context of the virtual case. This makes the process of editing a virtual patient cumbersome, and automated links between symptoms and conditions will be provided in the future to facilitate authoring. Fortunately, Freebase provides a list of associated symptoms for each condition it contains. The property used to link a condition to a symptom in Freebase has the URI `http://rdf.freebase.com/ns/medicine.disease.symptoms`. Figure 6.11 illustrates the mapping process that can be used to integrate symptoms and associated conditions found in Freebase to the SemVP model. This process was not implemented in SemVP, since if was not needed for the experimental purposes of this research(presented in chapters 7 and 8).

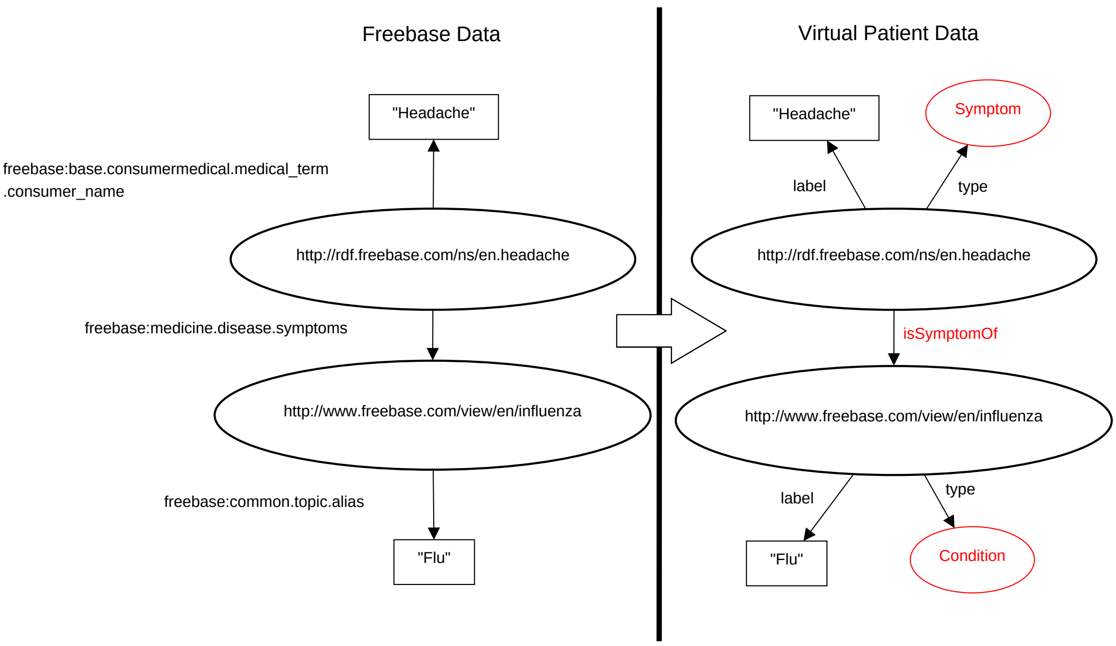


Figure 6.11: Mapping nodes from Freebase to classes from OpenGalen classes

The mapping process can be detailed as follows:

- Load triples linking conditions to symptoms from Freebase into SemVP,
- Map all symptoms from Freebase to the *Symptom* class. This process can occur automatically using a rule engine. The rule to apply would read as follows in natural language:
  `IF X` ***`http://rdf.freebase.com/ns/medicine.disease.symptoms`*** `Y, THEN X has type` ***`Symptom`*** `AND X has type` ***`Condition`*** `AND X` ***`isSymptomOf`*** `Y.`
  In other words, if X is considered to be a symptom in the Freebase knowledge base, it will also be a symptom of the same condition in SemVP, using the classes and properties provided by OpenGalen.
- Map properties specific to Freebase to simple RDF and RDFS properties.
  For instance, replace properties such as
  http://rdf.freebase.com/ns/base.consumermedical.medicsal_term.consumer_name or *http://rdf.freebase.com/ns/common.topic.alias* by the simpler, more generic property *label*.

A virtual case author can also link the patient's symptoms to a given condition manually, depending on the pedagogical goals of the case and the case author's expertise.

Using this structure, it is possible to identify relevant symptoms by querying all symptoms that are associated with the patient's condition (identified using the *hasDiagnosis* property). Since this thesis focuses mainly on feedback generation, this process was not implemented in SemVP. However, using Semantic Web technologies to facilitate virtual patient authoring is a valuable future research direction.

OpenGalen provides many classes and properties that were used to represent virtual patients in SemVP. The virtual patient's medical features are represented, and each one of these features can be a symptom or a risk factor. Each symptom or risk factor can be linked to a medical condition, using external data sources and expertise from a case author.

6.4 Representing Students' Actions in SemVP

6.4.1 Representing Interview Questions, Examinations and Investigations

Each of the patient's features can be observed by students during the clinical process. This is modelled in SemVP using OpenGalen and the *Observation* ontology design pattern.

After a student reads the virtual patient's presentation, the student can select three types of observations: interview questions, examinations, and lab tests. SemVP's interface contains a page with a list of each type of observations for students to choose from (see figures 6.1, 6.2, and 6.3). OpenGalen provides a class named *ClinicalAct*, and various subclasses of this class can be used to model each category of observation:

- *ConsultationAct* and its subclass *HistoryTakingAct* is used in SemVP's model to describe observations typically performed as part of the consultation process. This mainly includes history taking questions. *HistoryTakingAct* contains two subclasses designed to model two additional subtypes of questions: *FamilyHistoryTaking* and *PreviousPersonnalHistoryTaking*. No other subclass is available, so all history taking questions used for this study were created as instances of *HistoryTakingAct*.
- OpenGalen contains over a hundred subclasses under the *ClinicalAct* class, each designed to model a specific type of examination. Examinations modeled in OpenGalen include, among others, Abdominal examination (class *AbdominalExamination*, heart rate and breath sound examination (classes *HeartRateExamination* and *BreathSoundExamination*). *ExaminationAct* is a generic class contained in OpenGalen, which mean subclasses can be added to model specific examinations. If any relevant examination is missing from OpenGalen, it can be added to the virtual patient ontology as a subclass of *Examination*.
- *InvestigationAct* is a generic class designed to model investigation actions such as blood tests and other tests. It contains 92 subclasses, each modeling a specific type of test. It also contains a subclass called *LaboratoryExamination*. *LaboratoryExamination* can be used to model various types of lab tests, including but not limited to blood tests. In addition, three sub-classes are already available to model three specific types of blood tests: *FullBloodCount*, *BloodCoagulationTest* and *ProthrombinTimeTest*. Again, if any additional test is needed, it is straightforward to add it to the ontology.

These classes enable SemVP to model how each student accesses information describing the virtual patient. To do so, it is necessary to model the results of each of the student's observations. For this purpose, the "*Observation*" design pattern[6] is integrated to SemVP.
Ontology design patterns are generic class and property structures that can be reused to solve generic modeling problems (Gangemi, 2005). The "*Observation*" design pattern, for instance, is designed to model situations of observation, under a set of parameters. Figure 6.12 represents the general structure of this pattern.

[6] `http://ontologydesignpatterns.org/wiki/Submissions:Observation`

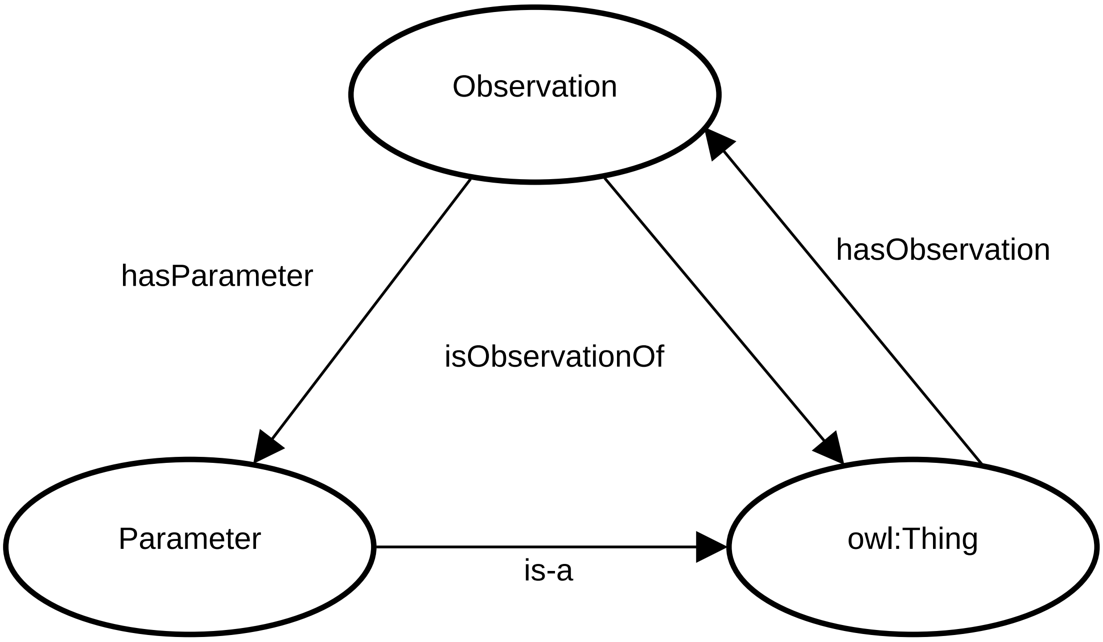


Figure 6.12: The Observation ontology design pattern (source: `http://ontologydesignpatterns.org/wiki/Submissions:Observation`)

Any action a student performs to assess the virtual patient and to define a diagnosis can be regarded as an observation of the patient. One can observe the patient as a whole, or observe a certain aspect of the patient. Since OpenGalen provides classes for both the body as a whole and for specific body parts and other aspects of the patient's physiology (see Section 6.4), it is then possible to link any aspect of the patient to one or several observations, using the *hasObservation* property provided in the *Observation* design pattern. The result of an observation is added to the virtual patient model using an additional class called *Result* and its associated property *hasResult*. This class has been created specifically for SemVP. A node representing an observation result can refer to a text, an image, a video file or a 3D animation. As an example, for the student to determine Ms. Matibunda's weight, two observations can be chosen: either ask a question (history taking act) or measure the weight using a scale. Figure 6.13 shows how this is modeled in SemVP.

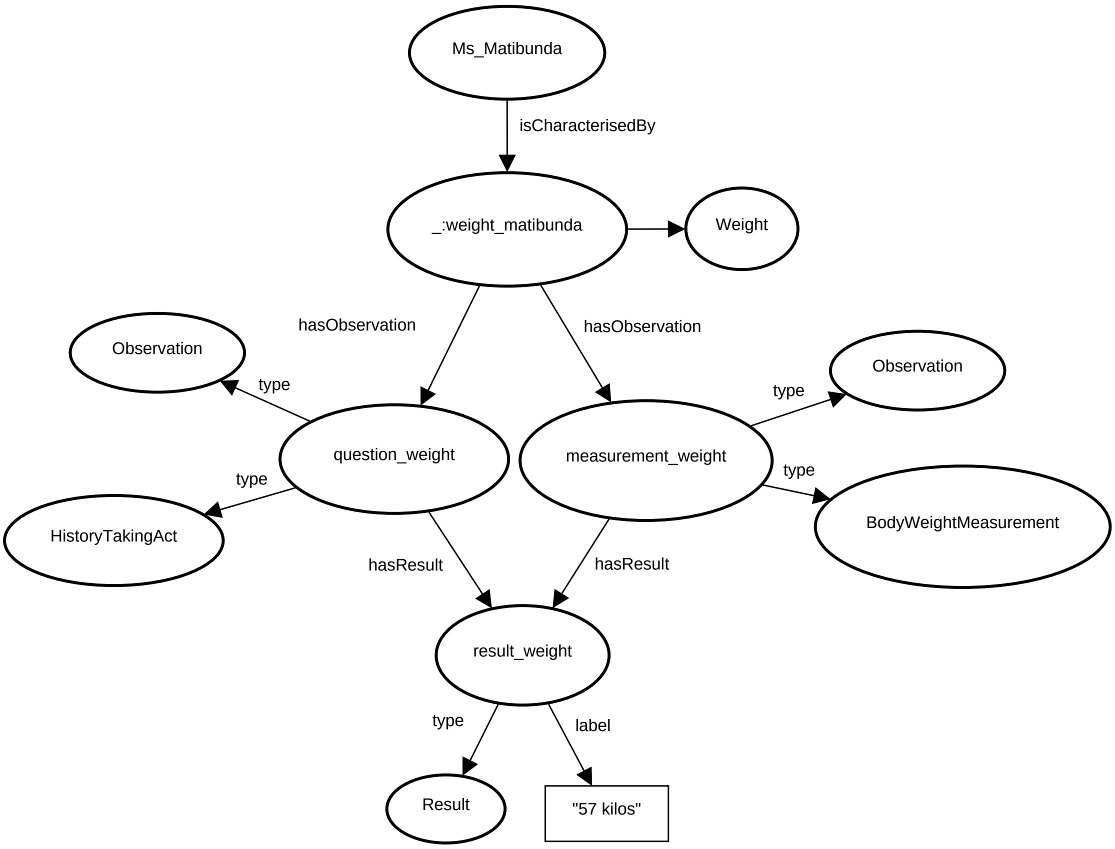


Figure 6.13: Example of Observation: determining Ms. Matibunda's weight either by asking a question or by measuring it.

Each possible action is represented by an RDF resource, belonging to the *Observation* class, and the patient's weight is represented by a resource belonging to the *BodyWeightMeasurement* class. The result of both these observations is a resource called *resul_weight*. This resource has a *label* annotation property showing the result in plain English.

To initiate a virtual case, a specific observation called the primary complaint is used. The primary complaint is generally represented by a resource belonging to the *HistoryTakingAct* class, and it represents the initial description of the patient's problem. The primary complaint is represented using an ad hoc class called *PrimaryComplaint*. Figure 6.14 shows how Ms. Matibunda's primary complaint is represented. In this case, she complains about altered sensation (paresthesia) and pain in her hands.

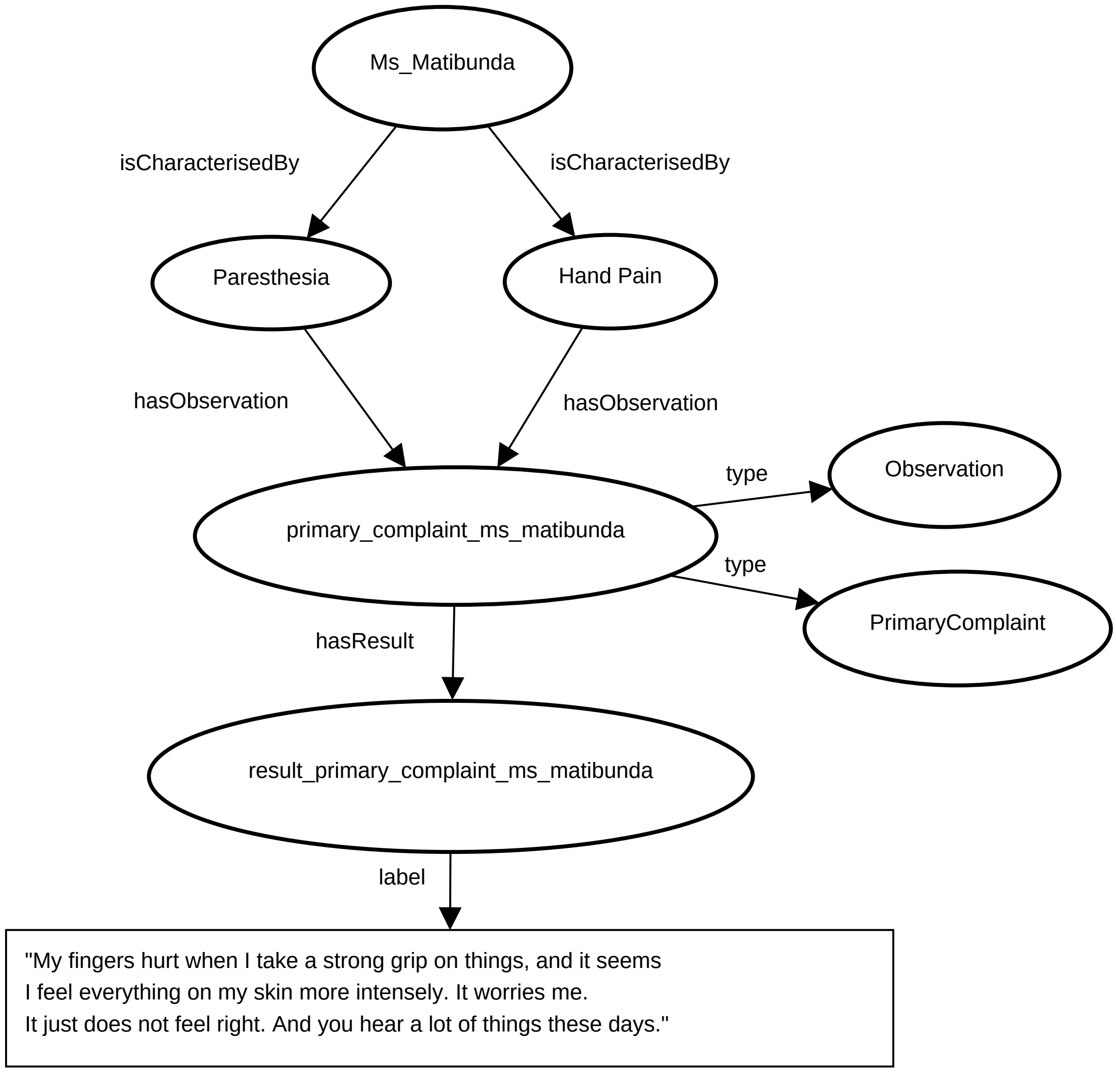


Figure 6.14: Representation of Ms. Matibunda's primary complaint.

6.4.2 A Student's Work Session: Recording and Retrieving Data About the Student's Actions

The structure described in the previous section implies that for a given virtual patient, any number of possible observations can be available to the student. Each observation is linked to one or several aspects of the virtual patient. However, to provide meaningful feedback to each student, it is necessary to know which observation has been chosen by each student, and when. During the course of an investigation, students will estimate the likelihood for each of their hypothesis, and adjust this estimation using new information. SemVP's model is designed to represent these evolving estimations.

To achieve this, SemVP is designed to represent *working sessions*. A session is a model representing all interactions performed by a student while investigating a virtual patient. Sessions are used to store data related to the student's actions, and to retrieve them when a student comes back to the virtual patient application after logging out. Each session involves one student and one virtual patient. A session contains all chosen questions, examinations, and lab tests chosen by a student while working on a virtual case. Figure 6.15 builds up on the example shown in figure 6.13 in the previous section. In this example, a student called *Student X* measures Ms. Matibunda's weight. To model this situation, a node of type *Access* is created, and is linked to the student's session and to the node representing the weight measurement. Additionally, a time stamp is linked to the *Access* node, in order to record the time of the student's choice.

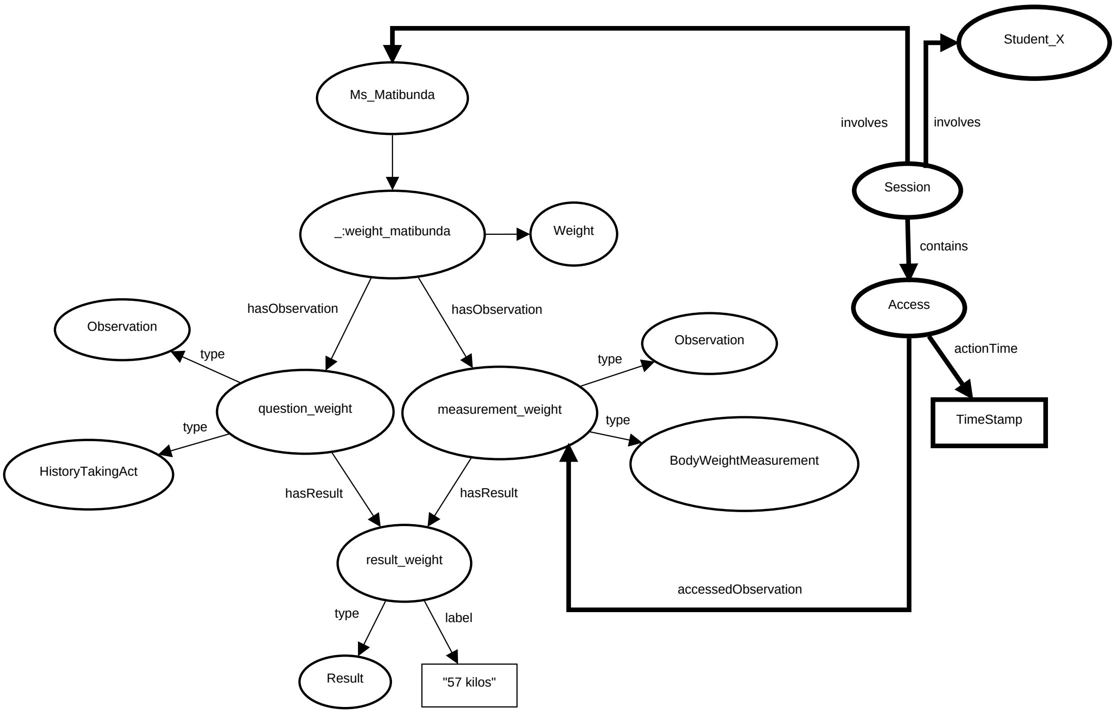


Figure 6.15: Example of student access to an observation

A session also contains the diagnoses that the student considered, associated with an estimated level of likelihood. Each time a student chooses a likely diagnosis, or changes the estimated likelihood of a previously proposed diagnosis, a new node of type "*DiagnosisProposal*" is created. Each *DiagnosisProposal* node is linked to a condition (the proposed diagnosis), a time stamp and a level of certainty (represented as an integer variable). The level of certainty has three possible values: 0 for a ruled out diagnosis, 1 for a diagnosis considered plausible, and 2 for a diagnosis considered likely.

Figure 6.16 shows an example of interaction model regarding diagnosis hypotheses. Katrin M. is affected by a fever, and the student proposes the flu, first as a likely diagnosis (likelihood level = 1). Later, the student decides to rule out this hypothesis based on new information. To do so, the student moves "Flu" from the "Likely" column to the "Unlikely" column in SemVP's interface (presented on figure 6.4).

In SemVP's underlying data model, this creates a new node linked to the student's session, and connected with a node representing the new level of certainty(0), and another node containing a new timestamp (the current time and date). When the student is ready to finish the case and submits a final diagnosis, the final diagnosis are identified in the SemVP's underlying model using the "hasFinalDiagnosis" property. In figure 6.16, the student has chosen "Flu" as the final diagnosis. This information is then used to generate feedback, as explained in section 6.5.

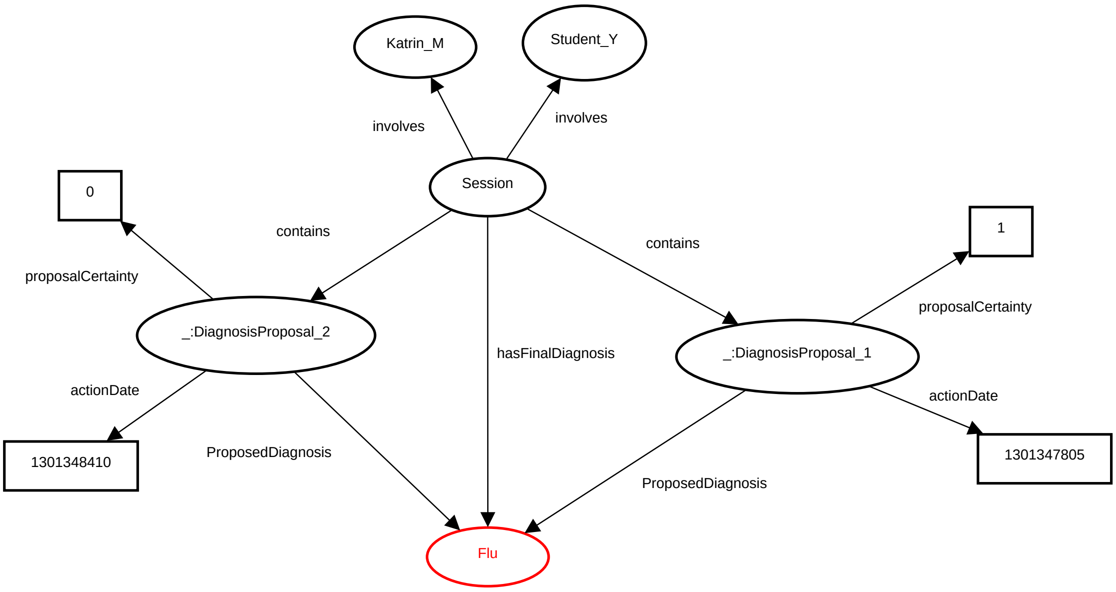


Figure 6.16: Example of student diagnosis choice model.

6.5 Generating Dynamic Feedback using Semantic Data

In SemVP, the *FeedbackGenerator* class generates the feedback displayed to students using data describing the patient and data describing the student's actions in the case. When a *FeedbackGenerator* object is instantiated, a *Session* object is used as a parameter to obtain the data from the student's work session. Thus, all data regarding the session can be retrieved, including the URIs of the patient, the student, and of each observation selected by the student.

Once these URIs are retrieved, it is possible to query the data and to generate feedback regarding the questions and examinations that the student has selected and the diagnoses proposed. This section describes how feedback is generated from this data in SemVP. The queries used in this section correspond to simple rules that generate new information using available data. These inferences could have been drawn using a rule-based reasoning engine(see section 3.4.3), but such a design would represent an important overhead, when SPARQL queries are enough to obtain the same result.

6.5.1 General Feedback about the Student's Final Diagnosis Choice

After gathering information to determine the correct diagnosis, the student selects a final diagnosis. Once this choice is made, the student is redirected to the final feedback page (see section 6.1). This page contains a short paragraph of general feedback, generated using the student's diagnosis choice and selected observation. Then, all observations selected by the students are listed in chronological order, with personalised feedback for each observations. The feedback provided for each observation changes dynamically depending on the observations chosen previously (see section 6.5.2).

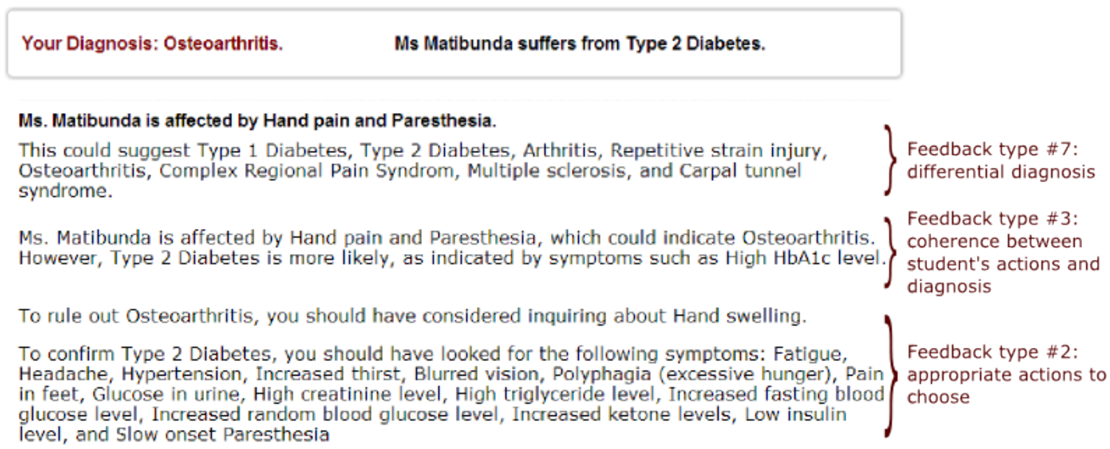


Figure 6.17: Example of dynamically generated feedback types #2, #3 and #7.

Figure 6.17 shows feedback generated by SemVP for a student who selected Osteoarthritis as a final diagnosis for Ms. Matibunda and used the following process to reach this diagnosis:

1. Ask question: "Can you describe the sensation in your hand a bit more?" (confirms a peripheral neuropathy),
2. Perform examination: "Examine Hands" (identifying the absence of swelling or stiffness on the hand),
3. Perform test: HbA1c test (strongly indicates diabetes),
4. Ask question: "Do you sometimes forget things more than you used to?".

First, the feedback lists various diagnoses that the primary complaint may indicate (feedback type #7, see Section 4.3). Then, the feedback highlights that a high HbA1c level is indicative of type 2 diabetes, helping the student to reflect on the interpretation of the results obtained from the HbA1c test (feedback type #3). Then the feedback indicates that the student should have considered looking for hand swelling to rule out osteoarthritis (a combination of feedback types #2 and #7). Finally, the feedback generates a list of suggested actions to take in order to confirm type 2 diabetes (feedback type #2).

The following paragraphs describe how this feedback is generated.
To start with, the feedback indicates if the diagnosis chosen by the student is correct. This is achieved by comparing the URI of the student's diagnosis to the URI of the correct diagnosis (see section 6.3 and figure 6.8 for details on how the correct diagnosis is represented in SemVP).
Then, more detailed feedback regarding the chosen diagnosis is provided. The first feedback provided is a list of conditions that could be consistent with the symptoms presented in the primary complaint (see section 6.3 for details of how the primary complaint is represented). To generate this list, the following SPARQL query is used for each symptom represented in the primary complaint:

```
SELECT ?PlausibleDiagnosis
WHERE {
     <Patient URI>   virtual_cases:isCharacterisedBy   <Symptom URI>.
     <Symptom URI>   opengalen:isSymptomOf             ?PlausibleDiagnosis.
     ?PlausibleDiagnosis    rdf:type                   opengalen:Condition.
}
```

Using this query, SemVP generates a feedback paragraph organised as follows:
"<*Patient's name*> is affected by <*first symptom in primary complaint*>, <*second symptom in primary complaint*>, [...], and <*last symptom in primary complaint*>.
This could indicate <*plausible diagnosis 1*>, <*plausible diagnosis 2*>, [...], and <*last plausible diagnosis*>."

This feedback corresponds to the feedback type #7 proposed in section 4.3: "A list of all diagnoses the student should have tested and ruled out, given the initial presentation of the patient". This part of the feedback is dynamically generated from the patient's data, which can be created by integrating various online sources (as demonstrated in section 6.3). It is designed to provide initial information about the patient's condition, which helps students to determine the relevance of their initial hypothesis.

The second paragraph of the feedback generated by SemVP is personalised using each student's choices. This paragraph indicates if the chosen diagnosis could be consistent with any of the patient's symptoms. The sentence generated is organised as follows:
"<*Patient's name*> is affected by <*symptoms that could be consistent with the student's chosen diagnosis*>, which could indicate <*student's chosen diagnosis*>.".

This feedback is displayed to help students reflect on the reasons that brought them to their final diagnosis.

The list of symptoms that could be consistent with the student's final diagnosis is generated using the following SPARQL query:

```
SELECT ?ChosenDiagnosisSymptom WHERE{
 <Patient URI>  virtual_cases:isCharacterisedBy  ?ChosenDiagnosisSymptom.
 ?ChosenDiagnosisSymptom  opengalen:isSymptomOf  <Chosen Diagnosis URI>.
}
```

If the student's diagnosis is incorrect, SemVP generates feedback explaining why the correct diagnosis is more likely. To do so, a query is used to return all symptoms identified by the student that are inconsistent with the student's diagnosis, but are associated to the correct diagnosis.

The sentence structure for this feedback is:

"However, <*correct diagnosis*> is more likely, as indicated by symptoms such as <*symptoms identified by students during the clinical process, which do not fit the student's diagnosis*>."

The list of identified symptoms that do not fit the student's diagnosis is generated using this SPARQL query:

```
SELECT ?Symptom WHERE{
   <Patient URI>  virtual_cases:isCharacterisedBy  ?Symptom.
   ?Symptom       observation:hasObservation       ?observation.

   ?Symptom                 opengalen:isSymptomOf      <Correct Diagnosis URI>.
   <Current Session URI>   virtual_cases:contains            ?actionAccess.
   ?actionAccess           virtual_cases:accessedObservation ?observation.
   FILTER NOT EXISTS{
        ?Symptom     opengalen:isSymptomOf    <Chosen Diagnosis URI>.
   }
}
```

First, this query selects all symptoms identified by the student and associated with the correct diagnosis. Then, all symptoms related to the student's diagnosis are filtered out. As a result, the query only returns symptoms that the student didn't take into account in the proposed diagnosis. The resulting feedback is designed to help students reflect on their interpretation of the symptoms they identify, by highlighting symptoms that are inconsistent with the diagnosis they proposed. **This generated feedback corresponds to feedback #3 ("Feedback telling the student if the chosen diagnosis is coherent with the results of the chosen interview questions and examinations").**

Another general feedback provided by SemVP highlights symptoms that students should have investigated to confirm their proposed diagnosis. This gives an indication to students about the symptoms they should have looked for before submitting their final diagnosis. The structure of the generated feedback sentence is a follows:
"To rule out <*chosen diagnosis*>, you should have considered inquiring about <*symptoms related to the chosen diagnosis, but that the student neglected to investigate*>".
A similar feedback is provided for risk factors linked to the chosen diagnosis that the student neglected to investigate. **This feedback corresponds to feedback type#2 from section 4.3: "A list of interview questions and examinations and tests the student should have chosen, and the justification (not appropriate, irrelevant, redundant, etc.)"**.

The feedback generated only provides a list of symptoms to investigate, and doesn't simply provide a list of observations to choose. This is designed to encourage students to reflect about the underlying causes of diagnosis error, and to think about which observations are required to investigate the relevant symptoms by themselves.

6.5.2 Feedback on Each Action Chosen by the Student

Figure 6.18 is an example of generated feedback designed to provide comments about each action selected by the student. For each action, the feedback indicates which conditions can be confirmed or newly suggested from the results or the patient's answer. Actions that confirm a diagnosis suggested by a previous action are highlighted in green (feedback types #2 and #4).

Your action	Feedback
1. Can you describe the sensation in your hand a bit more? *Well, it feels numb, and when I take a grip it's like there's many tiny needles in my hands.*	**Ms. Matibunda is affected by Paresthesia.** Well done, this could confirm a Peripheral Motor Neuropathy, Peripheral Sensory Neuropathy, and Peripheral Neuropathy. ✔ Well done, this could confirm Type 2 Diabetes, Multiple sclerosis, Carpal tunnel syndrome, and Osteoarthritis. ✔
2. Examine Hands *Hands appear normal, no rash, swelling or stiffness.*	**Ms. Matibunda is not affected by Skin rash, Stiffness, and Swelling.** Good choice, this could have confirmed Arthritis, Complex Regional Pain Syndrom, and Osteoarthritis. ★
3. HbA1c test *10% (normal range: 3.8-6.2%*	**Ms. Matibunda is affected by High HbA1c level.** Well done, this could confirm Type 2 Diabetes. ✔

Figure 6.18: Example of generated feedback for each action (1)

Figure 6.19 shows an alternative example of feedback, in a case where the student selected three irrelevant questions. Actions suggesting symptoms that are not consistent with any diagnoses proposed by the student or suggested by the visible symptoms are noted in red (feedback type #1).

Detailed description of your actions throughout the case

Your action	Feedback
1. Any back pain ? *No.*	**Ms. Matibunda is not affected by Back pain.** This might have suggested Kidney Cancer, Osteoporosis, and Bladder Cancer. However, this is not consistent with other symptoms.
2. Do you smoke? *I don't smoke.*	Tobacco smoking could have suggested Angina, Heart disease, Hypertension, Kidney Cancer, Lung Cancer, Osteoporosis, and Pancreatic cancer, but Ms. Matibunda is not affected by this risk factor.
3. Do you do a lot of manual work (sewing, painting, etc.) at work or at home? *No, I am retired and I don't really do that sort of activity. I prefer reading and Sudoku.*	Repetitive hand movements could have confirmed Carpal tunnel syndrome and Complex Regional Pain Syndrom, but Ms. Matibunda is not affected by this risk factor.

Figure 6.19: Example of generated feedback for each action (2)

6.5.2.1 Initialisation

To start the feedback generation algorithm in SemVP, plausible diagnoses are identified by inference. Plausible diagnoses are conditions that can be associated with the symptoms that the student can identify initially from the primary complaint. The primary complaint in the case of Ms. Matibunda, for instance, indicates two symptoms: paresthesia and pain in the hands. Paresthesia can indicate a variety of conditions such as diabetes, MS, carpal tunnel syndrome, osteoarthritis, or a peripheral neuropathy. Similarly, hand pain can indicate repetitive strain injury, diabetes, or arthritis. These links are formalised in RDF using the model described in section 6.3.3.2. Using this data, all these conditions can be considered as plausible diagnoses based on the information given in the primary complaint. To describe this in SemVP, the following SPARQL query is applied to SemVP's knowledge base to initiate the feedback generation process:

```
INSERT
```

```
{
   ?condition rdf:type virtual_cases:SuspectedDiagnosis.
}
WHERE
{
   ?symptom      observation:hasObservation        <Primary Complaint URI>.
   ?symptom      opengalen:isSymptomOf             ?condition.
   <Patient URI> virtual_cases:isCharacterisedBy   ?symptom.
   ?condition    rdf:type                          opengalen:Condition.
}
```

This query is equivalent to the rule:

```
IF a symptom observed in the primary complaint is a symptom of a condition,
THEN this condition is considered a plausible diagnosis.
```

Using this knowledge, it is then possible to determine, for each observation selected by the student, if this observation shows a symptom that is consistent with the information previously gathered. For instance, if the student enquires about chest pain in Ms. Matibunda's case, it is possible to determine that stiffness in the hands confirms arthritis, which was considered to be a plausible diagnosis because of the symptoms indicated in the primary complaint. Each new observations selected can confirm or infirm a plausible diagnosis. Each observations can also, by revealing a new symptom or risk factor to the student, suggest a new plausible diagnosis. This is described in the following section.

6.5.2.2 Feedback about the Relevance of Each Action

To enable students to reflect on their actions in detail, feedback is generated for each selected observation. The first sentence of the feedback indicates which symptoms or risk factors could have been identified through the observations selected.
To start with, the following query returns all symptoms affecting the patient that the currently selected observation can identify:

```
SELECT ?Symptom
WHERE {
    ?Symptom      observation:hasObservation         <Current Observation URI>.
    <Patient URI> virtual_cases:isCharacterisedBy    ?Symptom.
    ?Symptom      opengalen:isSymptomOf              ?anyCondition.
}
```

Using this query, SemVP returns the sentence: "<*Patient Name*> is affected by <*symptoms affecting the patient that can be identified through this observation*>".

The following query returns all symptoms that the current observation can identify, but that do *not* affect the patient:

```
SELECT ?Symptom
WHERE {
    ?Symptom      observation:hasObservation          <Current Observation URI>.
    <Patient URI> virtual_cases:isNotCharacterisedBy  ?Symptom.
    ?Symptom      opengalen:isSymptomOf               ?anyCondition.
}
```

Using this query, SemVP returns the sentences: "<*Patient Name*> is not affected by <*symptoms that are not affecting the patient, but that could be identified through this observations*".

These queries generate feedback giving indications about the relevance of each action chosen by the student. This can be related to feedback type #2 (appropriate actions).

Once all symptoms identified by an observation have been listed, it is possible to determine if the presence of each symptom confirms a plausible diagnosis. The following query is used to identify the plausible diagnoses associated to symptoms identified using an observation:

```
SELECT ?symptom ?condition
WHERE{
    ?symptom       observation:hasObservation          <Current Observation URI>.
    <Patient URI>  virtual_cases:isCharacterisedBy     ?symptom.
    ?symptom       opengalen:isSymptomOf               ?condition.
    ?condition     rdf:type         virtual_cases:SuspectedDiagnosis.
}
```

The sentence generated by SemVP using this query has the following structure:"Well done, this could confirm <*diagnoses confirmed by the symptoms identified*>".

The absence of a symptom can also give students indication about a plausible condition. Using the following query, plausible conditions that could have been confirmed by an observation are identified:

```
SELECT ?symptom ?condition
WHERE{
```

```
    ?symptom       observation:hasObservation         <Current Observation URI>.
    <Patient URI>  virtual_cases:isNotCharacterisedBy  ?symptom.
    ?symptom       opengalen:isSymptomOf               ?condition.
    ?condition     rdf:type        virtual_cases:SuspectedDiagnosis.
}
```

Such observations are considered relevant, since they help students to rule out certain diagnosis hypotheses. The sentence generated by SemVP using this query has the following structure: "Well done, this could have confirmed <*diagnoses that could have been confirmed by the presence of a symptom*>".

Finally, an observation can reveal a symptom or risk factor suggesting a new diagnosis. For instance, if the student enquires about blurry vision in Ms. Matibunda's case, this symptom might suggest Multiple Sclerosis (MS). Since Ms. Matibunda is indeed affected by blurry vision, the condition MS (and other conditions associated to blurry vision) is added to the list of plausible diagnoses using the following query:

```
INSERT
{
   ?condition rdf:type virtual_cases:SuspectedDiagnosis.
}
WHERE
{
   ?symptom      observation:hasObservation        <Current Observation URI>.
   ?symptom      opengalen:isSymptomOf             ?condition.
   <Patient URI> virtual_cases:isCharacterisedBy   ?symptom.
   ?condition    rdf:type                          opengalen:Condition
}
```

The feedback generated from this is: "<*Symptom identified* > could suggest <*new conditions suggested by the symptom*>".

A similar query is also used to underline risk factors linked to other possible diagnoses. The resulting feedback is related to the differential diagnosis that students formulate using the available information (feedback type #7). It also relates to the relevance of each action to confirm or rule out certain possible diagnoses (feedback type #2).

Using this algorithm, SemVP takes into account all observations previously selected by the student to generate feedback on the current observation. For instance, if the student enquired about memory loss after asking about blurry vision, the feedback would indicate "memory loss would have *confirmed* MS". This is because MS has been added as a plausible diagnosis while generating feedback for the previous observations.

However, if the student enquires about memory loss without previously asking about blurry vision, the feedback would only indicate "memory loss could *suggest* MS".
Using these queries, SemVP can generate feedback related to the sequence of students' actions (feedback type #4), and to the relevance of each chosen action (feedback type #1 and #2).

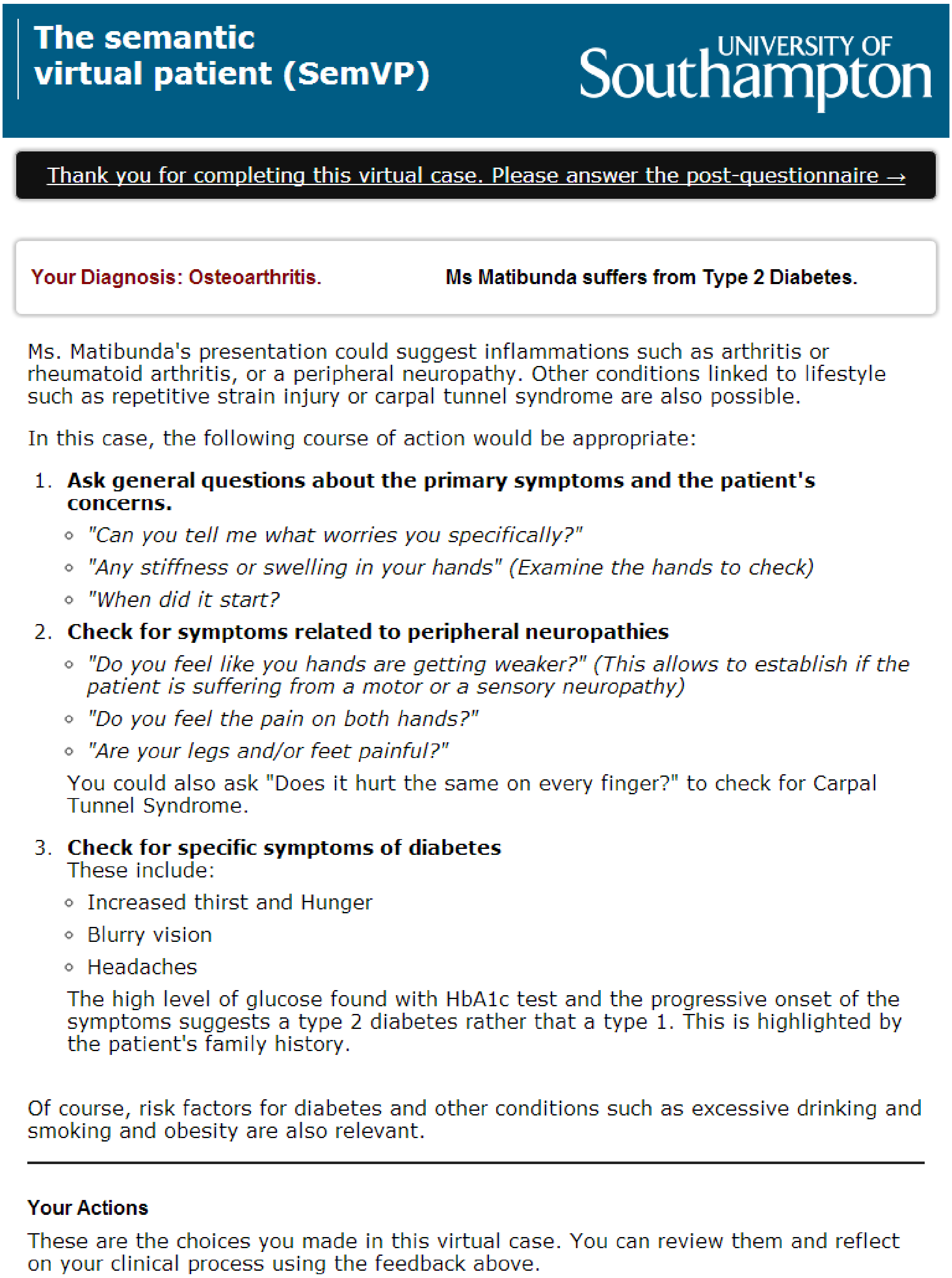
6.6 Static Feedback for Experimental Comparison

The experiment presented in Chapter 7 was designed to determine how valuable students find SemVP's automatic feedback compared to a generic, manual feedback. To make this comparison, a static feedback text was written based on Ms. Matibunda's case (see Section 5.4.1). This feedback was delivered to a randomised control group of students, and rated on Likert Scales using various criteria (Chapter 7 contains the detailed experimental design).

The statically authored feedback provided to the control group was designed to highlight the important steps of the clinical process in Ms. Matibunda's case:

1. Initial assessment about primary symptoms,
2. Assessment of symptoms related to peripheral neuropathies (most probable diagnosis given the nature of the patient's hand pain),
3. Narrowing down on diabetes by checking for specific symptoms such as increased thirst and blurry vision,
4. Confirming the diagnosis using lab tests.

This static feedback was designed to provide general guidelines about the case, and to help students reflect on their reasoning by comparing it with the process proposed. It was designed using the analysis performed on Ms. Matibunda's case during the preliminary study (see Chapter 5), and also using interviews conducted with clinicians (described in section 7.2).
The feedback provided is displayed as shown in figure 6.20:



The semantic virtual patient (SemVP)

UNIVERSITY OF Southampton

Thank you for completing this virtual case. Please answer the post-questionnaire →

Your Diagnosis: Osteoarthritis. **Ms Matibunda suffers from Type 2 Diabetes.**

Ms. Matibunda's presentation could suggest inflammations such as arthritis or rheumatoid arthritis, or a peripheral neuropathy. Other conditions linked to lifestyle such as repetitive strain injury or carpal tunnel syndrome are also possible.

In this case, the following course of action would be appropriate:

1. **Ask general questions about the primary symptoms and the patient's concerns.**
   - *"Can you tell me what worries you specifically?"*
   - *"Any stiffness or swelling in your hands" (Examine the hands to check)*
   - *"When did it start?*
2. **Check for symptoms related to peripheral neuropathies**
   - *"Do you feel like you hands are getting weaker?" (This allows to establish if the patient is suffering from a motor or a sensory neuropathy)*
   - *"Do you feel the pain on both hands?"*
   - *"Are your legs and/or feet painful?"*

   You could also ask "Does it hurt the same on every finger?" to check for Carpal Tunnel Syndrome.
3. **Check for specific symptoms of diabetes**
   These include:
   - Increased thirst and Hunger
   - Blurry vision
   - Headaches

   The high level of glucose found with HbA1c test and the progressive onset of the symptoms suggests a type 2 diabetes rather that a type 1. This is highlighted by the patient's family history.

Of course, risk factors for diabetes and other conditions such as excessive drinking and smoking and obesity are also relevant.

Your Actions

These are the choices you made in this virtual case. You can review them and reflect on your clinical process using the feedback above.

Figure 6.20: Static feedback delivered to the randomised control group (see chapter 7)

After this feedback, SemVP provides a chronological list of all actions selected by the student, enabling the student to revise their clinical process in the light of the information provided in the static feedback.

Semantic web technologies allow the representation of complex knowledge, and are also designed to facilitate the reuse of existing data, as explained in Chapter 3. SemVP leverages these features. It is based on a semantic data model that represents both the virtual patient and each students' actions. This model uses pre-existing data sources such as the OpenGalen ontology, the Observation design pattern, and the Freebase knowledge base. Using this model, SemVP generates a paragraph containing feedback about each student's diagnosis choices. In this feedback, the coherence (or absence thereof) between the student's choices and the final diagnosis is highlighted. Relevant observations that the student has neglected are also highlighted. Following this general feedback, a comment is generated for each of the student's chosen observations. The experiment presented in Chapter 7 provides evidence to show that semantic web technologies can facilitate the automatic generation of rich and individualised feedback in virtual patients.

Chapter 7

Study Methodology

A mixed method study was conducted in order to evaluate the benefits of the feedback generated from SemVP's underlying semantic model. This study was conducted with students who started their clinical assignments, i.e. students in year 3, year 4 and final year. Students from two universities were involved in this study: the University of Southampton (UoS) and the Karolinska Institute (KI) in Stockholm.

The main objectives of the study were to determine if students see benefits in the automatic and personalised feedback generated by SemVP and if they consider that automated feedback improves their understanding of the virtual case more than static feedback.
The secondary objective was to identify any variation between the answers provided by students from different year groups (year 3, 4 and final year) and from different schools.

The opportunity to conduct the study in two schools was taken in order to establish that SemVP's model can provide useful feedback across different cultural and educational contexts. The UoS Faculty of Medicine and the KI are inherently different in cultural context. However, their curricula are similar, in that students learn the basic medical sciences in the first two years, and start their clinical training in year 3. Therefore, a comparison between schools can help to establish if differences in culture have an impact on students' understanding of feedback, and on the benefits provided by SemVP. A comparison by year group was conducted in order to determine if differences in students' experience have an influence on their requirements about feedback and their understanding of feedback. It also helps to determine if SemVP provides similar benefits across year groups.

A mixed method approach was used to account for the richness of the students' experience and the richness of the clinical reasoning process. It is difficult to understand why students make certain decisions in a clinical case using only quantitative data, due to the number of factors involved: previous clinical experience, variations in medical school training, students' intuition, etc.. As a result, determining the most useful types of feedback is also difficult using an exclusively quantitative approach. This warrants a methodology where quantitative and qualitative data are used together and complement each other to reach a clear picture of the student's thought process (Malterud, 2001).

7.1 Case Design and Validation with Clinicians

For the purpose of this experiment, Ms. Matibunda's case (described in chapter 5) was used. This case was chosen because the initial presentation is not obvious, thus the case requires students to conduct a thorough investigation, which offers many possible data points for analysis.
The original Medbiquitous case was enhanced with additional questions, examinations and lab tests, using the results from the preliminary interviews with students presented in Section 5.4.1. This enabled students to have a wider range of interactions to choose from. In total, 26 history taking questions, 5 examinations and 11 lab tests were provided in SemVP. The complete data set representing the case is presented on table A.1, in appendix A.

After the implementation of Ms. Matibunda's case in SemVP, 6 clinicians (4 general practitioners, a neurologist and a pediatrician) were interviewed to validate the case and ensure that the interactions used in SemVP were easy to understand. Each clinician was invited to use SemVP and determine the correct diagnosis using SemVP. While interacting with the case, they were invited to "think aloud" and explain their choices of interactions. Clinicians were interviewed as they were going through the case. This semi-structured discussion was the first part of the interview.

After going through the case, clinicians were invited to discuss feedback in virtual patients, and the role of virtual patients in medical education in general. The objective of these unstructured discussions was to determine clinicians' perception of SemVP, and of virtual patients in general, in particular in terms of feedback for clinical reasoning development.

These interviews were transcribed by listing all the actions performed by each clinician in chronological order. Next to each action, the corresponding comments explaining why the action was chosen were transcribed. Finally, clinicians' comments about the case and about feedback in general were transcribed. To analyse these transcripts, the most common actions selected were grouped by frequency and by theme.

The emerging pattern in these interviews was to start with general and open questions about the patient's symptoms and to address the patient's concerns early on. Then, clinicians established that the symptoms were characteristic of a peripheral neuropathy, and asked more specific questions to define a more precise diagnosis. Finally, clinicians checked for specific symptoms of diabetes using further questions and lab tests. This pattern appears clearly in the static feedback presented in Section 6.6.

7.2 Study Design

After this validation of the case by clinicians, medical students in Year 3, Year 4 and Final Year were invited to participate in this experimental evaluation of SemVP. Students from both the UoS and from the KI evaluated SemVP. Students from the UoS took part in the study in March and April 2012, and students from KI participated in the study in May and June 2012. Figure 7.1 is an overview of the study design.

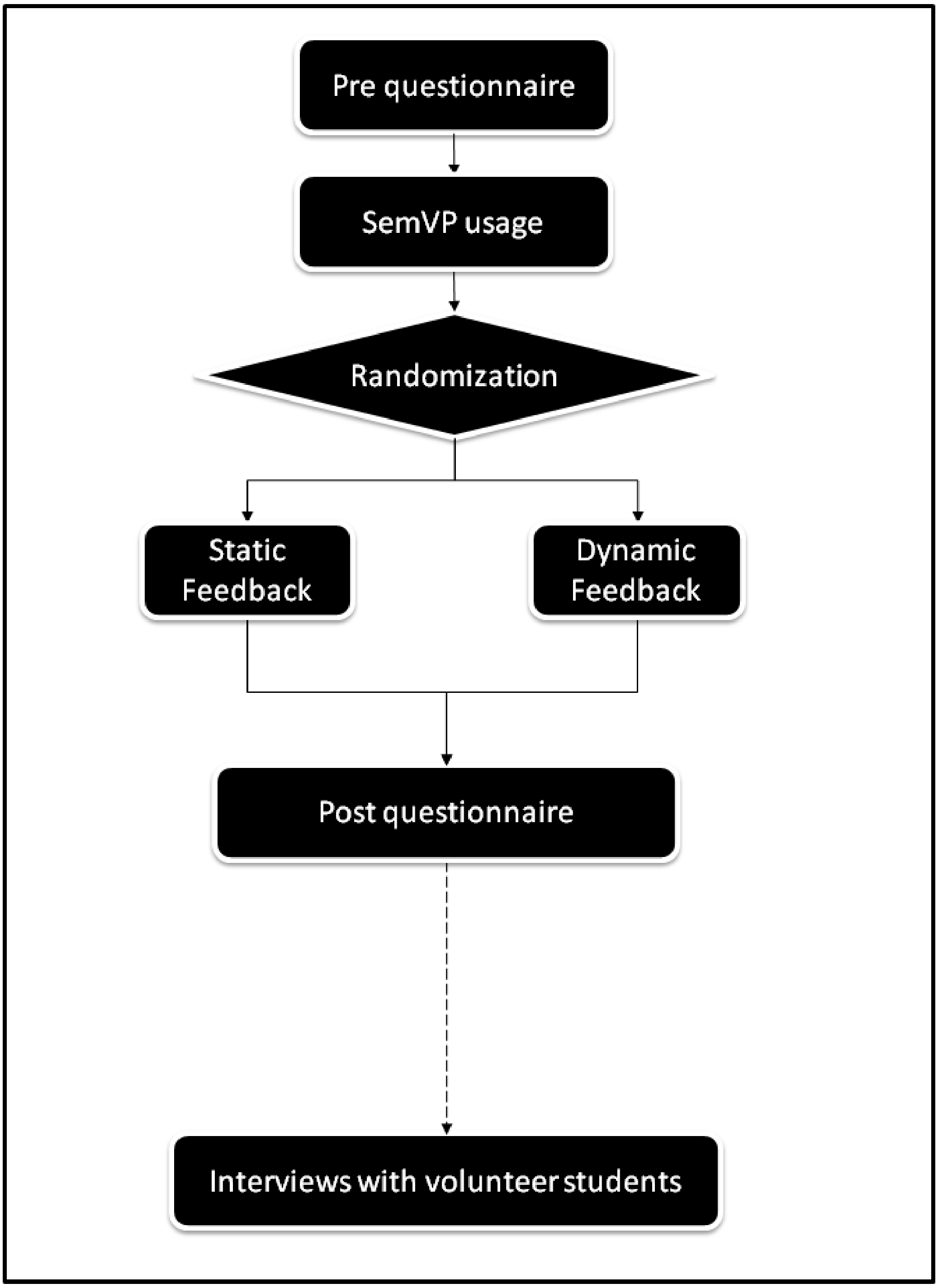


Figure 7.1: Study design

The first step of the experiment was an online pre-questionnaire about feedback in virtual patients and virtual patients in general (see details in Section 7.2.1). All participating students answered this questionnaire.

Then, all participants interacted with SemVP as it is described in chapter 6. Participants' actions within SemVP were also recorded using the model described in Section 6.4. After this interaction, participants submitted a final diagnosis, and received feedback from SemVP. In order to compare the automatic feedback generated (described in Section 6.5) with static feedback (presented in Section 6.6), a random control group of participants received static feedback after using SemVP, and the rest of the participants received dynamic feedback. The feedback method was assigned using a block randomisation algorithm, stratified by year group and using blocks of four, as follows:

1. Create a list of 10 variables containing 5 variables with a value of 0 ("false"), and 5 with a value of 1 ("true"). This ensures that 4 consecutive participants are divided in two equal groups of 5.
2. Shuffle the list randomly, so there is no way to determine in which order the variables are sorted in.
3. Every time a participant in this year group clicks to obtain feedback, assign static feedback if the variable has a value of 0, and dynamic feedback if the variable has a value of 1. Use the next variable on the list for the following participant in the same year group, and the next for the participant after that, until the end of the list is reached.
4. Repeat the operation for the next 4 participants in the same year group.

Using this method, participants were randomly divided in roughly equal groups for each year group in each school.

After receiving feedback from SemVP, all participating students answered a post-questionnaire described in Section 7.2.2. To avoid external disruptions and allow participants to focus on the task, this evaluation was done in the presence of the researcher, using the Health Services Library computer room in Southampton and a lecture theater with university-provided laptops at KI.

Participants were also invited to volunteer for an interview after completing the case and answering the pre and post-questionnaires (see Section 7.2.3). During the interviews, volunteer students were invited to discuss the questions, examinations and lab tests they chose, and the rationale behind each choice, including the diagnoses the suspected. They were also asked about the feedback they received from SemVP, and to discuss virtual patients in general, in particular the role of feedback in virtual patients.
All questionnaires and interview protocols were validated by the UoS Faculty of Medicine ethics committee. They were also reviewed by the clinicians interviewed before the study, and by two medical students in Year 4 and 5.

7.2.1 Pre-Questionnaire

The pre-questionnaire used in this study, very similar to the preliminary survey described in chapter 5, is divided into two sections. Section 1 is focused on participants' year group and previous experience with virtual patients (see figure 7.2), and section 2 is devoted to students' expectations and requirements for a new virtual patient system (see figure 7.3).

Pre Questionnaire

Your user name

We will provide a user name for your. Please type it here.

Part I – Previous Experience with Virtual Patients

What year of study are you in?

- Year 2
- Year 3
- Year 4
- Final year

Have you used virtual patients previously during your medical education?

- Yes
- No

In what context have you used virtual patients before (multiple choices possible)?

- As a self-paced learning and revision tool
- As a compulsory assessment in a module
- As a group study tool

Any other comment about how you have used virtual patients in the past?

Feedback provided by virtual patients used previously

Given your learning needs and your previous experience with virtual patients, rate the following statements:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The feedback provided by virtual patient systems I used in the past was useful.					
The feedback provided by virtual patient systems I used in the past was relevant to my learning needs at the time.					

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Self-study virtual patients would be a useful tool to improve my decision-making and diagnosis skills during my clinical training.					

Figure 7.2: Pre-questionnaire part 1: previous experience with virtual patients

Information was gathered regarding students' year group, the context of their previous use of virtual patients, and how useful they found the feedback delivered in the virtual patients they used in the past. Data gathered from section 1 was used to determine the commonalities and differences in participants' opinions of virtual patients and their experience with them. It was also used to compare students' answers between year groups.

Part II – Expectations and Requirements for a New Virtual Patient System

Given you current experience and level of study, which part of the diagnostic process do you find most difficult? (tick one answer)
(In these statements, the general term "test" refers to any interview question, examination and lab test that might be relevant to the clinical process:)

- ○ Knowing which conditions to test for given the patient's initial presentation
- ○ Knowing the relevant history questions to ask and the relevant examinations to perform given the patient's initial presentation
- ○ Interpreting the information obtained through tests
- ○ Adjusting the differential diagnosis using the patient's answer to each question and the result of each test

Any other comment about your difficulties with the clinical process?

What kind of feedback would you find most useful now, in a self-learning and self-assessment virtual patient system?

	Useless	Not Very Useful	Neutral	Useful	Very Useful
A list of interview questions and examinations I should not have chosen, and the justification (not appropriate, irrelevant, redundant, etc.).	○	○	○	○	○
A list of interview questions and examinations I should have chosen, and the justification (type of disease to consider, related symptom(s) to check, possible conditions to rule out, etc.).	○	○	○	○	○
Feedback about the order in which I chose specific actions.	○	○	○	○	○
Feedback telling me if the diagnosis I chose is coherent with respect to the results of the interview questions and examinations I performed.	○	○	○	○	○
If I chose an inadequate action, a narrative description of the consequences to the patient, if applicable.	○	○	○	○	○
A list of all diagnoses I should confirm or rule out, given the initial presentation of the patient.	○	○	○	○	○
A sequence of the "ideal" history taking and examination process that an expert would use, with the rationale for each step.	○	○	○	○	○

Submit

Figure 7.3: Pre-questionnaire part 2: Expectations and requirements for a new virtual patient system

Section 2 is divided in two groups of questions. The first two questions are related to the student's difficulties in the diagnostic process. In the last part of the pre-questionnaire, the students were invited to rate the 7 types of feedback described in Section 4.3 on a 5-step Likert scale ranging from "Useless" to "Very Useful".

7.2.2 Post Questionnaire

The objective of the post-questionnaire was to enable students to rate the feedback they received from SemVP. A comparison was then possible between the ratings given to static feedback and the rating given to dynamic feedback. Participant were invited to rate the feedback received using the following statements, using a 5-step Likert scale ranging from “Strongly disagree” to “Strongly agree”:

- “This feedback improved my understanding of the case.”
- “This feedback changed my initial assumptions about the case.”
- “This feedback helped me to understand my errors and evaluate my choices.”
- “This feedback was adapted to my current level of expertise.”
- “Practicing other cases with the same type of feedback will be beneficial to me.”

Participant were also asked to rate the usefulness of each of the 7 types of feedback a second time (this was also asked in the pre-questionnaire). This was done in order to identify any change in student’s opinions about each feedback type after receiving feedback from SemVP (see figure 7.4).

Post-Questionnaire

Please rate the following statements regarding the feedback you received:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
This feedback improved my understanding of the case.	○	○	○	○	○
This feedback changed my initial assumptions about the case.	○	○	○	○	○
This feedback helped me to understand my errors and evaluate my choices.	○	○	○	○	○
This feedback was adapted to my current level of expertise.	○	○	○	○	○
Practicing other cases with the same type of feedback will be beneficial to me.	○	○	○	○	○

After receiving feedback in this virtual patient, please re-evaluate the following types of feedback:

	Useless	Not Very Useful	Neutral	Useful	Very Useful
A list of interview questions and examinations I should not have chosen, and the justification (not appropriate, irrelevant, redundant, etc.).	○	○	○	○	○
A list of interview questions and examinations I should have chosen, and the justification (type of disease to consider, related symptom(s) to check, possible conditions to rule out, etc.).	○	○	○	○	○
Feedback about the order in which I chose specific actions.	○	○	○	○	○
Feedback telling me if the diagnosis I chose is coherent with respect to the results of the interview questions and examinations I performed.	○	○	○	○	○
If I chose an inadequate action, a narrative description of the consequences to the patient, if applicable.	○	○	○	○	○
A list of all diagnoses I should confirm or rule out, given the initial presentation of the patient.	○	○	○	○	○
A sequence of the "ideal" history taking and examination process that an expert would use, with the rationale for each step.	○	○	○	○	○

Submit

Figure 7.4: Post-questionnaire

7.2.3 Interviews

After using SemVP, volunteer participants were interviewed in order to understand the choices they made while solving Ms. Matibunda's case. Each interview was based on the chronological list of actions that the participant had chosen. While reviewing this list, the participant was asked the following questions for each action:

- "Did you have a diagnosis hypothesis at this point? If you did, what made you think about this diagnosis? What would you do to rule out or confirm this hypothesis?"
- "Why did you choose this action? Which outcome were you expecting to obtain from this action? Were you trying to confirm or rule out a given hypothesis?"
- "In retrospect, what do you think of this decision? What would you have done differently if you used this virtual patient again?"

Then, participants were asked open questions regarding feedback and virtual patients in general, such as: "What are you looking for in virtual patients? What is your opinion on feedback in virtual patient and in clinical practice in general? How do you use virtual patients? How would you use a virtual patient like SemVP if you had access to it on a regular basis? Any other comment?", etc..

Each possible observation in SemVP (interview questions, examinations, lab tests) was attributed a unique identification number. Then, each interview was transcribed as follows:

1. All actions chosen by the interviewee were listed in chronological order using the data logged in SemVP (see Section 6.4),
2. Comments formulated by the interviewee regarding each chosen observation were transcribed,
3. General comments about virtual patients and feedback regarding virtual patients were also transcribed.

7.3 Analysis

The following analysis was performed:

- A thematic analysis of the interviews, to explore interviewees' perceptions about the feedback delivered by SemVP,
- An statistical analysis of the ratings given by participants for each feedback type in the pre and post-questionnaires, to establish if the year group, school, and provided feedback (static or dynamic) have an effect on participants' ratings,
- A statistical analysis of the rating participants give for each criteria on the post-questionnaire, comparing participants who received static feedback with those who received dynamic feedback.

To perform the interview analysis, the observations most frequently selected by interviewees from the UoS and interviewees from the KI were identified. Then, interviewees' comments were coded and grouped, in order to identify recurrent patterns and themes, in particular comments related to each type of feedback identified in chapter 4. General comments related to the clinical process, medical good practice, hypothesis generating and testing, and feedback were also explored and analysed. Comments related to each of these topics were coded by sub-topic, and compared by year group and school.

These qualitative results were analysed in conjunction with quantitative data from participants' rating of the feedback delivered by SemVP. The ratings for each of the 5 criteria in the post-questionnaire were compared between participants who received statically authored feedback and those who received dynamic feedback.

Since the distribution of the result was difficult to determine, Mann-Whitney U tests were initially used to compare the mean rating given for each criteria. Indeed, non-parametric tests are generally considered more reliable when the distribution of the data is unknown. However, Norman (2010) showed that parametric tests are actually robust in practice for Likert scale analysis, even with non-normal distributions and relatively small samples. Thus, unpaired t-tests were also performed. In cases were t-tests returned similar results to Mann-Whitney U tests, parametric tests were considered reliable, therefore parametric tests based on the Univariate General Linear Model (GLM) were used to determine the effect of year group and school on participants' answers.
A similar approach was used to compare the mean rating provided for each feedback type in the pre and post-questionnaires. The results were initially compared using Wilcoxon signed-rank tests (which is used to compare means between paired variables). Then, paired t-tests were also performed. When t-tests and Wilcoxon signed-rank tests returned similar results, parametric tests were used to determine the effect of year group and school on participants' answers.

All statistical tests were performed using SPSS v.20.0, with a 95% confidence level. The results were reviewed by a medical statistician at the UoS Faculty of Medicine.

Chapter 8

Results of the Study: The Benefits of Automatic Feedback Generated by SemVP

Twenty medical students from the UoS and 45 students from the KI participated in the study. Participants randomly received either static feedback or dynamic feedback, as described in Chapter 7.

Table 8.1 shows how participants were distributed across schools, year groups, and type of feedback received[1].

	UoS Faculty of Medicine		Karolinska Institute		**Total**
	Static Feedback	Dynamic Feedback	Static Feedback	Dynamic Feedback	
Year 3	4	3	14	26	47
Year 4	2	1	3	2	8
Final Year	6	4	0	0	10
Total	12	8	17	28	65

Table 8.1: Distribution of participants to the experiment

Thirteen of these participants volunteered to be interviewed afterwards(7 from the UoS and 6 from the KI). Interviewed participants referred to feeback types #1, #2, #3 and #7 directly or indirectly when discussing their thought process. Analysis of the quantitative data confirms the importance of these feedback types. Interviewees' comments referring to each type of feedback confirm the benefits of SemVP's generated feedback.

[1]The distribution of Year 3 students at KI shows that students who received dynamic feedback outnumber those who received static feedback by 12. This was due to an unforseen reset of the web server by the UoS technicians, which deleted the file used to count students for the randomisation algorithm. However, this did not alter the statistical significance of the results.

Throughout the interviews and questionnaires, participants also demonstrated a consistency in their understanding of feedback and in their requirements regarding feedback.

Pariticpants also underlined the importance of personalised feedback, mentioning their appreciation of SemVP's adaptability to their actions. The quantitative data confirms this: SemVP's automatic and personalised feedback is generally rated higher than the alternative static feedback in terms of improved understanding of the case, changed initial assumptions, and improved understanding of each participant's mistakes for self-evaluation.

8.1 Most useful Types of Feedback Identified by Students

During the interviews, interviewed students made comments that can be directly or indirectly related to several feedback types discussed in Chapter 4. Three main categories of comments have been identified:

- Comments related to mistakes interviewees made or potential mistakes that they identified,
- Comments related to important actions to perform in the case, in order to reach the correct diagnosis,
- Comments related to interviewees' expectations for feedback in general.

An analysis of these comments shows the benefits that SemVP's automatic feedback delivers for each type, compared to a generic and static feedback.

8.1.1 Feedback Regarding Students' Proposed Diagnoses

One of the most cited feedback type is type #7 (differential diagnosis). Interviewees described several possible diagnosis mistakes which can be related to feedback type #7 as they discussed their progression through the case. While analysing their comments about these mistakes, themes related to feedback #3 (coherence diagnosis-actions) emerged.

Three main types of mistakes were identified by interviewed participants:

- focusing prematurely on one diagnosis, excluding other possibilities,
- disregarding relevant information,
- failing to implement important or relevant actions.

Premature focus on one diagnosis was seen in practice, after two participants (both from the UoS) failed to reach the correct diagnosis, and discussed their mistakes using the information provided in the feedback. One of these students (in Final Year) stated, after receiving automatic feedback: "In hindsight, I was a bit too focused on MS, and I didn't take into consideration other factors like her age, which would have led me away from it.". The student showed an ability to reflect on the clinical process itself, admitting: "I've latched on it (MS) and tried to fit things to MS rather than trying to look what it was, and reach a differential. I went about it the wrong way.". This interview showed that premature focus on an initial diagnosis can lead to disregarding relevant information. This student also noted that the feedback provided was helpful for reflection about the clinical process, and highlighted a recurrent tendency in his clinical process: "[The feedback] was really useful, it highlighted that I've done what I always do. I need to work on that".

The other student who didn't reach the correct diagnosis (in Year 3) exhibited the same type of mistake, focusing early on osteoarthritis because of the patient's age and symptoms. This student also showed the same tendency to fit the received information to the initial diagnosis instead of using additional information to adapt the differential diagnosis. For example, the student enquired about potential weakness in the patient's hands, and commented "Osteoporosis could limit the movements of the hands, making it seem like it's getting weaker". Whilst most other interviewed participants used this question to establish if the patient was suffering from a motor neuropathy or a sensory neuropathy (see Section 8.1.2), this participant used the question only to confirm the initial diagnosis.
In another example, this student enquired about swelling in the patient's hands, but commented in retrospect: "The answer to this question should probably have changed my initial diagnosis, but I didn't think about it.". Ten out of the 13 interviewed students ruled out joint problems after noting the absence of swelling or stiffness on the patient's hands (see Section 8.1.2).
Contrary to the first student in Final Year, this student did not explicitly mention how the feedback provided helped to reflect on her clinical process. However, she did exhibit an ability to reflect on her own process, as demonstrated by the above-mentioned comments. This may be due to the fact that this student was only in Year 3, with a limited clinical experience, thus unable to relate the diagnosis process performed in SemVP to past experiences in a clinical setting.

Five participants from all schools and age groups mentioned the possibility of making similar mistakes, and how it could lead to an inability to reach the correct diagnosis or even look for the most relevant pieces of information because of an initial bias towards a given hypothesis. The mistakes described by these five students also relate to the coherence between the actions taken by students using the SemVP and the diagnosis they chose. Students' hypothesis guide their choice of actions, as well as their interpretation of the information they gather. Knowing if they gathered the right information and if they interpreted it correctly, regardless of the initial hypotheses they might have formulated, is a valuable piece of information to these participants. Thus, the mistakes identified also relate to feedback type #3 (coherence diagnosis-actions).

Three participants from the KI also explicitly commented on feedback type #7, stating that having an initial differential diagnosis was useful, in order to check their initial assumptions: "It's nice to have a differential at the beginning, that way you can say: these are all the things that I ruled out.". These participants all received dynamic feedback, and the opinion they expressed was that knowing which condition they should have looked for in hindsight is valuable to them, and helped them reflect on their own initial hypotheses.

8.1.2 Feedback about Appropriate Choices to Make

Comments related to feedback type #2 (appropriate actions) were also very frequently made.
First, interview questions, examinations or lab tests were considered important if they helped participants to confirm or rule out a likely condition. For instance, a Year 4 student from the UoS asked the patient about shortness of breath, indicating: "[..] for a woman that age I thought it was important to check for cardiac problems". Nine participants, across all groups, checked for arthritis or other joint problems, given the patient's age and symptoms. In all cases but one (see previous section), joint problems were ruled out based on to the absence of swelling or stiffness in the hands. All but one participant from the UoS also mentioned that they suspected a peripheral neuropathy, and confirmed this before narrowing their search to diabetes. Participants from the KI, however, talked about diabetes without previously mentioning a peripheral neuropathy. It is difficult to know where this difference comes from from this study alone, and one can only speculate that this is due to variations in the participants' clinical learning environments, the teaching methods and content of their medical teaching programmes or the type of clinical experience students have access to.

Important actions were also used to differentiate between several initial hypotheses and identify the most probable diagnosis by eliminating others. Across all groups, students asked questions to differentiate between a local trauma on the patient's hands (such as injury) vs. systemic condition. Participants from the UoS in all age groups also differentiated between sensory and motor neuropathy. These differentiations occurred early in the process, allowing them to narrow down their diagnosis to diabetes afterwards. Table 8.2 summarises the most frequently cited diagnoses hypotheses, and the questions used to confirm them, rule them out or differentiate them.

Stated Goal	*Associated Actions*
Differentiate between a local or a systemic condition	• "Any stiffness in your hands?" (question #1) • "Any swelling in your hands?" (question #2) • "Can you describe the sensation in your hand a bit more?" (question #4) • "Are your feet painful?" (question #6) • "How long have you had this problem?" (question #9) • "Is it painful on both hands?" (question #14)
Differentiate between a sensory and a motor neuropathy	• "Do you feel like your hands are getting weaker?"
Rule out arthritis	• "Any stiffness in your hands?" (question #1) • "Any swelling in your hands?" (question #2) • "Can you describe the sensation in your hand a bit more?" (question #4) • "Is it painful on both hands?" (question #14)
Confirm that the patient suffers from a peripheral neuropathy	• "Can you describe the sensation in your hand a bit more?" (question #4) • "Are your feet painful?" (question #6) • "Is it painful on both hands?" (question #14)

Table 8.2: Frequently asked questions and frequently associated goals

Two participants (one from the KI and one from the UoS) mentioned failure to implement important or relevant actions as a possible clinical mistake. Sometimes relevant actions that can be missed, as explained by the participant from the UoS: "I could have asked about weight loss, which is common in type 2 diabetese, but I didn't think about it at the time.".

The static feedback created for the experiment provides information describing the clinical process, and showing students an appropriate course of action to take in Ms. Matibunda's case. Differentiation between motor and sensory neuropathy was mentioned, as well as confirming diabetes.

One student from the KI remarked: "You can become a 'mindless clicker' while using [virtual patients], rather than analysing the information that's given". Another explained: "In VPs it's so easy to click everything just to be on the safe side.". Since SemVP provides information about each action chosen, students can use this information to reflect on the relevance of each of their actions. Another interviewee, talking about the feedback provided by SemVP, explained: "It would have been great to have more feedback about questions that are really unnecessary and redundant.". Indeed, some questions can be considered less relevant in some cases, and sometimes it is useful to indicate this in the feedback (feedback type #1).

However, students had contrasting views about this type of feedback. Three students from both schools in years 4 and 5 remarked that very few actions are actually completely irrelevant, even though some are more relevant than others. One student in year 4 from the UoS emphasised the need to help students focus, i.e. to look for specific things related to the initial complain. But this student also noted that it is important to have a broader view and "keep an open mind".

Six participants from all groups emphasised the need to choose the most relevant actions first, or otherwise indicated that the order in which questions are chosen is important to them (feedback type #4). Several factors were cited to explain the importance of actions' order, which highlights the complexity involved in delivering automatic feedback about the order of actions:

- The most important questions have to be asked early on, in order to reach the diagnosis efficiently (and possibly to maintain the patient's trust). For instance, open questions were highlighted as important because they can lead to "early wins" [sic]. Then questions regarding the symptoms directly can help to obtain a clearer picture of the problem.
- Students need to adjust their differential diagnosis and their actions to the information they receive.

A Final Year student from the UoS expressed doubts about feedback regarding the order of students' actions. The reason given was that "in real situations your next question is led by what is being said to you.". Thus, it is difficult to generate useful sequence-related feedback, since this type of feedback is very context-sensitive.

SemVP provides feedback that changes depending on sequence of selected actions, as described in Section 6.5. Even though the feedback does not specifically give information about the order of actions, the process chosen by students is taken into account to generate feedback. This isn't possible to achieve with static feedback. Only a static text detailing the most appropriate course of action (according to the feedback author) can be provided. This type of "roadmap" is actually similar to feedback #5 (expert process), but participants did not make any direct or indirect reference to this type of feedback.

8.1.3 Students' Rating of Each Feedback Type

Quantitative data analysis highlight the importance of feedback types #2, #3, and #7, which are consistently rated higher than other feedback types, both before and after using SemVP, as showed on figure 8.1. Friedman tests, Kendall's W tests and repeated measures ANOVA showed a very high consistency in ranking for each feedback type, both before and after using SemVP ($p<0.001$).

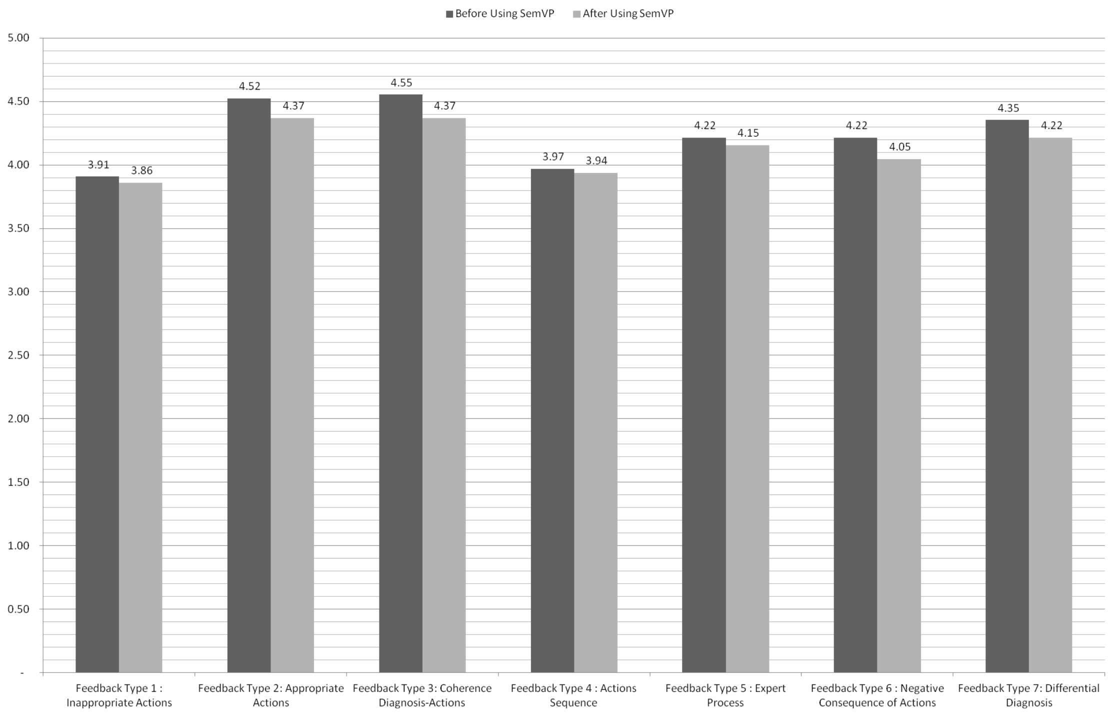


Figure 8.1: Mean ratings of perceived usefulness for each feedback type overall, before and after using SemVP

Interview data described in Section 8.1 showed that interviewed participants repeatedly mentioned the same feedback types (mainly #2, #3 and #7), across schools and across year groups. The comments they made about each type of feedback are also similar across groups. However, participants in year 3 did not exhibit an ability to reflect on their own process, as demonstrated in the example described in Section 8.1.1.

Statistical tests were used to determine if participants' views about each type of feedback changed after using SemVP. Wilcoxon signed-rank tests and paired t-tests were used to determine if student's rated each feedback type significantly lower or higher after using SemVP. Non-parametric tests (Wilcoxon signed-rank) yielded very similar results to the parametric paired t-tests. Therefore, parametric tests were considered reliable for these results (see Section 7.3).

The usefulness of each feedback type was consistently rated slightly lower after using SemVP (see table 8.3). However, both Wilcoxon signed-rank tests and paired t-tests show that this difference is only significant for feedback type #3 (p:.028). Additionally, the differences in ranking are very small (the confidence interval of the difference for feedback #3 is between .021 and .348). One can only speculate what these differences could be attributed to. A number of factors may have had an influence, such as the content of the feedback provided by SemVP during the experiment, a change in the participants' perception due to their actions in the case, or simply a general tiredness towards the end of the experiment. The size of the sample may also have influenced the results.

	Before	*After*	*Diff.*	*Sig. (Wilco.)*	*Sig.(t-test)*
Feedback Type 1: Inappropriate Actions	3.91	3.86	0.1	.65	.658
***Feedback Type 2*: Appropriate Actions**	**4.52**	**4.34**	**0.2**	**.10**	*.096*
***Feedback Type 3*: Coherence Diagnosis-Actions**	**4.55**	**4.37**	**0.2**	**.028**	**.027**
Feedback Type 4: Actions Sequence.	3.97	3.94	0.03	.834	.788
Feedback Type 5: Expert Process	4.22	4.15	0.1	.63	.559
Feedback Type 6: Negative Consequence of Actions.	4.22	4.05	0.2	.057	.055
***Feedback Type 7*: Differential Diagnosis**	**4.35**	**4.22**	**0.1**	**.274**	**.228**

Table 8.3: Mean Score fore Each Feedback Type, Before and After Using SemVP.

ANOVA tests (Univariate GLM) were used to determine if participants from different schools and different year groups rated each feedback type differently. A significant difference was found between year groups for feedback type #1 (p: 0.016). Indeed, participants in Year 4 and Final Year generally rated feedback type #1 lower than participants in year 3, especially before using SemVP (see figure B.3 in appendix B). This is also reflected in the interview data (see Section 8.1.2). In other feedback types, the year group or the school attended by participants did not have any significant influence on their rating of each feedback type. However, since participants in Year 3 were largely over-represented in the sample, a different sample may show more significant differences between year groups.

8.2 Students' Perception of SemVP's Dynamic Feedback Compared to Static Feedback: the Crucial Role of Personalisation in Feedback

One of the main themes emerging from the interviews is the importance of personalisation in the feedback. Nine interviewees from all groups indicated that they enjoyed receiving feedback related to their own actions, which helped them understand how they performed. They mentioned appreciating the information associated to each of their actions, and the self-evaluation that this feature allowed. One student from the KI explained this: "I think the feedback was good. There was an explanation about why the question was the right choice. **It made me evaluate my choices for myself**.". A student who received static feedback acknowledged its benefits (differential diagnosis), but remarked that a more personalised feedback was lacking, because there wasn't any indication about the student's performance: "The feedback was useful, as it gave information about the differential diagnosis. It doesn't tell me anything about how I did specifically, though.". Overall, interviewed participants wanted to know how well they performed individually, to see why each of their actions was relevant (or not), and to evaluate their choices throughout the case. Feedback types #2, #3 and #7 provide valuable and personalised information for students to do so, as demonstrated in Section 8.1.

Interestingly, two students (both from Stockholm) expressed opposing views about which part of the automatic feedback was the most useful. One of them appreciated the first, general part of the feedback (regarding the submitted diagnosis and related symptoms, see Section 6.5.1), but did not read the feedback related to each action because "it seemed unnecessary, it was too long[..]". The other student took the opposite view and focused on feedback related to each action chosen (see Section 6.5.2), explaining: "The first general feedback had a lot of text, so I didn't bother and I went straight to the feedback on my questions. I was more interested in knowing if my questions were

good.”.
A common element between these comments is the length of the feedback. This is probably due to the fact that feedback was delivered entirely at the end of the case. Other modes of feedback delivery need to be explored to avoid overwhelming students with long sections of textual feedback (see Section 9.3).

This emphasis on personalised feedback is confirmed by ratings from the post-questionnaire. Figure 8.2 and 8.3 show that participants who received dynamic feedback from SemVP generally agree more with each statement proposed. The difference in rating is more pronounced for the first three statements, which are directly related to the benefits of feedback in terms of reasoning improvement and self-reflection (see figure 8.2).

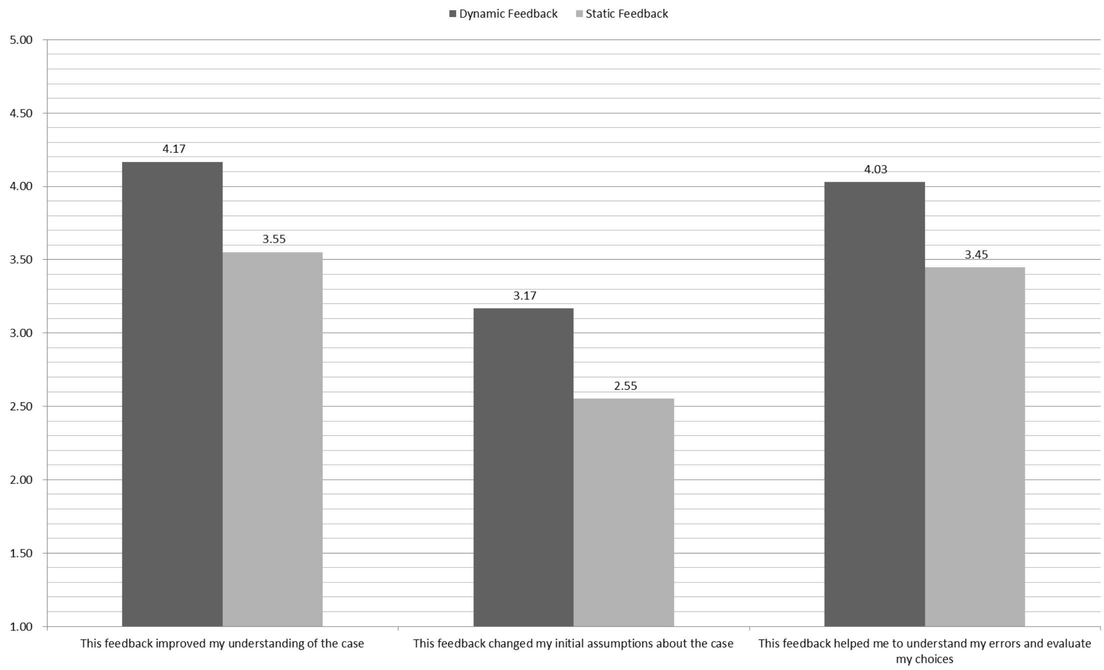


Figure 8.2: Mean ratings for the first three statements, for static and dynamic feedback

The difference in mean rating is smaller for the last two statements, which relate to how feedback was adapted to students' educational needs.

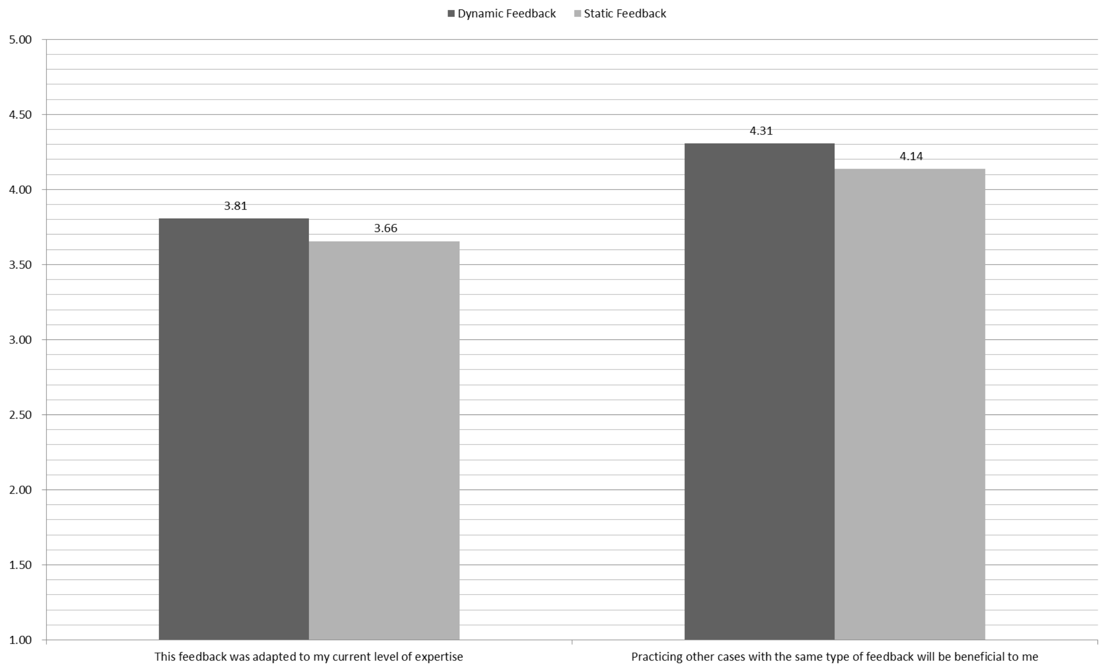


Figure 8.3: Mean ratings for the last two statements, for static and dynamic feedback

Table 8.4 shows that the mean difference in rating between participants who received static feedback and those who received dynamic feedback is significant for the first three statements. Both Mann-Whitney U tests and independent samples t-tests confirm the statistical significance of this result. Both tests show no significant differences in ratings for the last two statements. This shows that participants considered both static and dynamic feedback to be adapted to their level of expertise, and that participants see benefits in using feedback provided in virtual patients for practice in clinical reasoning. Figure 8.3 shows that participants in general indicated a neutral or even negative opinion (for the static feedback) about how much the feedback provided to them changed their initial assumptions about the case. This is confirmed by the interviews, where most interviewees reached the correct diagnosis using similar observations than those used by clinicians and in a similar order, and did not express any surprise when receiving the feedback.

Criteria	Mean Difference	Sig. (M.-Wit.)	Sig. (t-test)	95% Confidence Interval	
				Lower	Upper
"This feedback improved my understanding of the case."	.61	.001	.001	.237	.993
"This feedback changed my initial assumptions about the case."	.61	.015	.010	.152	1.078
"This feedback helped me to understand my errors and evaluate my choices."	.58	.003	.001	.200	.959
"This feedback was adapted to my current level of expertise."	.15	.393	.48	.277	.578
"Practicing other cases with the same type of feedback will be beneficial to me."	.17	.165	.36	.181	.530

Table 8.4: Mann-Whitney U tests and T-Tests results for all criteria: difference in mean rating between students who received static feedback and those who received dynamic feedback (results produced using SPSS).

Since parametric tests showed similar results to the non-parametric equivalent, parametric tests were used to establish if year groups and medical schools had an effect on participants' ratings for each statement. ANOVA tests (Univariate GLM) showed no significant effect between year group and school on participants' ratings, despite some variations in means (see figures B.5, B.6, B.7 and B.8 in appendix B).
However, since students in Year 3 were largely over-represented in the sample, a different sample may show different results.

8.3 Discussion

The static feedback delivered by SemVP during the experiment provided indications about likely conditions that could affect the patient, such as arthritis or rheumatoid arthritis, or a peripheral neuropathy (see Section 6.6). It also provided information about how to differentiate between the most likely hypotheses in order to narrow down the diagnosis to diabetes.
However, qualitative and quantitative data gathered for this study confirms that the feedback generated by SemVP provides more personalised information, based on data

describing the patient's information and participant's actions. In particular, SemVP generates a list of symptoms affecting the patient that are coherent with the student's final diagnosis, indicating why the student's choice might be appropriate in each case. Additionally, a list of symptoms that are inconsistent with the student's chosen diagnosis is generated, which helps students to reflect on relevant information that they may have missed, or important actions that they should have chosen.
Thus, SemVP provides similar information to that contained in the static feedback, but in a way that is adapted to the each student's choices. Examples from the interviews show that students have the ability to reflect on their mistakes when they focus on one diagnosis and fail to consider other possibilities. Given this ability, personalised feedback can help each student to reflect on their own mistakes, without the need to look for the information most relevant to them in a generic feedback. The data confirms that students value the dynamic feedback generated by SemVP.

SemVP provides feedback directly based on each student's action, which is difficult to achieve (even impossible in some cases) when authoring feedback statically. Students exhibit an ability to identify their mistakes and reflect on their actions, in particular in Final Year. They know how to use virtual patients, and understand how virtual patients can be useful as training tools. They want to use virtual patients for self-evaluation, to know how well they perform and why. SemVP's personalised feedback allows them to do so. As a consequence, participants appreciated the personalised feedback delivered by SemVP regarding each of their actions. Indeed, SemVP enables students to reflect on each of their choices, thus producing feedback as a "cycle" (as opposed to feedback simply delivered as information) and allowing them to elaborate and adjust clinical schema using SemVP (see Chapter 4).

Using static feedback, students need to compare their actions to the proposed process written in the static feedback. This could add a cognitive load that may hinder self-evaluation. By contrast, the dynamic feedback delivered by SemVP provides feedback directly related to each choice that students make. Since the feedback is displayed directly next to each action, students only have to use the information provided for their own actions, which allows them to focus directly on their own thought process (see Section 8.2). Since SemVP can generate feedback at any moment in the diagnosis process, students may use the information at various times to adjust their decisions based on the feedback provided. This was also mentioned by a participant from the KI: "[..]some people might work through the whole thing and get something completely irrelevant. Having indications as you go would be useful, especially in year 3.". This is not achievable using static feedback, unless the clinical process is separated into predetermined stages (for instance, separating history taking, examination, lab tests, and management plan), and ad-hoc feedback is written for each stage.

SemVP delivers information for each of the student's choices, allowing students to evaluate the relevance of each action for themselves, depending on the hypotheses that can be confirmed or ruled out by the action and the actions' place in the sequence of students' choices. This cannot be achieved easily with static feedback, due to the large number of possibly irrelevant actions that students can choose from. The fact that each action is more or less relevant depending on the context also increases the complexity involved in this type of feedback, thus making static feedback for this very difficult to produce.

Participants also placed a high value on personalisation in feedback, and indicated that this aspect of SemVP's generated feedback was very valuable to them. Interview data shows that students want to use feedback to evaluate their performance and understand their mistakes. As a consequence, participants rated SemVP's dynamic feedback higher that static feedback on 3 out of 5 criteria, even though both feedbacks provided essentially the same information and are rated similarly in terms of value to medical education and adequacy to participants' level. The results of this study show that students appreciate the ability to reflect on each of their choices using personalised feedback, and SemVP provides a practical solution for this.

With few exceptions, the results are consistent across year groups and schools. However, since this study was focused mainly on Year 3 students, who constituted a large proportion of the investigated sample, a different sample may reveal some differences between year groups.

Chapter 9

Conclusion and Future Work

This chapter summarises the research presented in this thesis, and synthesises its the contributions to the field of virtual patients. Future research opportunities in this field are then proposed.

The main focus of this thesis was the automated generation of personalised feedback in virtual patients, using semantic web technologies.

A classification of seven different types of feedback that help students improve their clinical reasoning skills was proposed. During a preliminary study, this classification was evaluated by medical students using an online questionnaire. They consistently considered two types of feedback to be the most useful: feedback regarding the correct course of action to take (feedback type #2) and feedback indicating the conditions to look for in each virtual case (type #7). This was consistent across year groups 3 to 5.

SemVP, a virtual patient system based on Semantic Web technologies, was designed and implemented. It runs using a semantic model that represents the patient's symptoms and conditions, as well as the actions chosen by each student. Using this information, SemVP generates individual feedback that provides information to each student about their final diagnosis and their performance in terms of clinical reasoning. SemVP leverages preexisting data available on the web, such as the OpenGalen ontology, the Observation ontology design pattern, and the Freebase knowledge base.

In a final experimental study, SemVP was evaluated by students from the University of Southampton Faculty of Medicine and from the Karolinska Institute in Stockholm, using a virtual case created from an example available on the eVip repository. This evaluation was conducted using a mixed-method approach. A questionnaire designed to evaluate students' opinion of the feedback delivered by SemVP was completed by 65 students. Semi-structured interviews were also conducted with a subgroup of 13 volunteer students. SemVP's automatic feedback was compared to a static feedback using a randomised control group, stratified by year group and medical school. The participants consistently indicated that SemVP's automatic feedback improved their understanding of the virtual case, changed their initial assumptions about the case, and helped them understand their mistakes and evaluate their choices. Evidence for this was also found in the interview data, and these results appeared consistently across schools and year groups. During the interviews, participants also highlighted the importance of personalisation in feedback, and the value of measuring their individual performance. They demonstrated an ability to reflect independently on their performance and to identify their mistakes. SemVP's individualised feedback was showed to be effective in supporting these activities.

9.1 Contributions

9.1.1 A Classification of Feedback Types for Clinical Reasoning

Feedback is at the heart of virtual patients' role in medical education, since it allows students to adjust their reasoning process based on an evaluation of their performances. Feedback enables students to create and adjust clinical schemes, ie. prototypical patient stories that experienced clinicians use to apply their scientific knowledge in a clinical context. Personalised feedback is especially useful for this purpose, as it allows students to reach a higher standard of performance starting from their own mistakes (see Chapter 4).

Seven feedback types were identified from the literature. These types of feedback were highlighted because they help students identify reasoning mistakes such as faulty hypothesis triggering, faulty context, faulty information processing and faulty verification. Feedback types from this classification relate to these mistakes, and helps students reflect on their choices. They are listed as follow:

- *Feedback Type 1* (Inappropriate Actions): "A list of interview questions, examinations and tests the student should *NOT* have chosen, and the justification (not appropriate, irrelevant, redundant, etc.)".
- *Feedback Type 2* (Appropriate Actions): "A list of the interview questions and examinations students *should* have chosen, and the justification (type of disease to consider, related symptom to check, possible conditions to rule out, etc.)".
- *Feedback Type 3* (Coherence Diagnosis-Actions): "If the diagnosis is wrong, feedback telling the student if the chosen diagnosis is still coherent with the results of the chosen interview questions and examinations".
- *Feedback Type 4* (Actions Sequence): "Feedback about the order in which the student performed specific actions".

- *Feedback Type 5* (Expert Process): "A sequence of the "ideal" history taking and examination process that an expert would use, with the rational for each step".
- *Feedback Type 6* (Negative Consequences of Actions): "If the student chooses an inadequate action, a narrative description of the consequences on the patient, if applicable".
- *Feedback Type 7* (Differential Diagnosis): "A list of all diagnoses the student should have tested and ruled out, given the initial presentation of the patient".

The feedback types identified as most useful during the preliminary study were feedback type #2, and #7, as described in Chapter 5. In the final study presented in Chapters 7 and 8, participants identified feedback #2, #3 and #7 as most useful. They rated the usefulness of each feedback type consistently before and after using SemVP. They also mentioned feedback types #2, #3 and #7 explicitly or indirectly during the interviews. These results were consistent across both medical schools and year groups.

9.1.2 A Semantic Model of Virtual Patients

Presented in Chapter 6, SemVP is based on a detailed semantic model of virtual patients, which uses preexisting semantic data from the OpenGalen ontology, the Observation ontology design pattern and the Freeebase knowledge base. This model enables SemVP to represent a virtual patient in a computer-processable manner. SemVP can represent many social, physiological and anatomical features. Each feature can be accessed using observations, which are represented as questions, examinations or lab tests. Observations results can be represented as text, picture, videos, or any type of web-retrievable resource. The choices made by each student in their interactions with the virtual patient are also stored in a RDF (see Chapter 3 for more details on RDF and other Semantic Web languages) using this semantic model.

9.1.2.1 Automatic and Individualised Feedback in Virtual Patients

Combining data related to the patient and data describing each student' actions in the system, SemVP generates automatic feedback for each individual student, based on their choices. This feedback gives information about each student's diagnosis choices, and highlights the coherence (or absence thereof) between the student's choices and the final diagnosis. A comment relative to each of the student's chosen observations is also provided. When evaluating SemVP, participants consistently highlighted the importance of personalised feedback, and demonstrated the ability to reflect on their choices and to analyse their path in the virtual patient system. They consistently indicated that the automatic feedback delivered by SemVP helped them understand the virtual case better, challenged their assumptions, and helped them evaluate their mistakes. This was shown through a quantitative survey, comparing SemVP's automatic feedback with static feedback containing equivalent information. Supporting evidence was also found from the interviews with 14 students.

9.2 Publications

The research described in this thesis was presented in three peer-reviewed conferences. Below is a summary of these publications:

1. *Semantic virtual patients: using semantic web technology to improve virtual patients for medical education*, 3rd international conference on web science, Koblenz, Germany (14 - 17 June 2011)
   This paper presents the benefits of the Semantic Web over other technological approaches for virtual patients. It presents the limitations of existing virtual patient systems in terms of feedback, and presents how the Semantic Web helps to alleviate these limitations. SemVP's model was also described.

2. *Modelling Virtual Patients and Generating Feedback using Semantic Web Technologies*, ASME Annual Scientific Meeting, Edinburgh (13 - 15 July 2011)
   This paper presents an overview of virtual patients and their uses in medical education. It also describes the benefits of personalised feedback, and the difficulties associated with its authoring. Finally, SemVP's model and its feedback generation algorithms were presented.

3. *Automatic Feedback Generation in Virtual Patients Using Semantic Web Technologies*, 2011 International Computer Assisted Assessment (CAA) Conference, Southampton, UK (05 - 06 July 2011)
   Through a case study involving a student solving a virtual case, this paper described SemVP's feedback generation algorithms in detail, and presents its educational benefits.

9.3 Future Work

The use of semantic web technologies for virtual patients design can lead to valuable research, from both technological and pedagogical perspectives.

Technology-related research in this field may involve extending SemVP's data model. The model could be extended for several purposes:

- To increase the level of detail in the virtual patient model, leading to a more precise and exhaustive representation of the patient's physiology and anatomy,
- To integrated probabilistic components to the SemVP model, in order to generate more complex and detailed feedback using Bayesian statistics.

A flexible model such as SemVP's can benefit from the increasingly detailed and complex data available to describe and model the human body. The complexity involved in using this type of data within virtual patients presents interesting challenges in the fields of algorithms, artificial intelligence, and expert systems.

In its current state, SemVP uses a simple, binary model linking symptoms to corresponding conditions, using data from external knowledge bases directly. However, in real clinical situations, the link between a particular symptom or finding and a diagnosis is not usually this simple, but depends highly on the context. As formalised by Bayes' theorem (Molina, 1931), the probability of a given diagnosis depends on previously obtained results, a concept that students are taught to use in clinical reasoning early in the curriculum. Using probabilistic models as an extension of SemVP implies many challenges in artificial intelligence, bio-informatics and algorithms.

SemVP also presents potential for research in medical education and pedagogy. Such research could consist of:

- The comparison of various interaction designs and modes of feedback delivery enabled by SemVP's model,
- The extension of the semantic model presented in this work to emotional states, in order to provide feedback on students' attitudes and behaviours,
- The effects of SemVP and its use on students' learning in actual teaching and learning settings.

Various modes of feedback delivery enaled by SemVP could be compared from a pedagogical point of view. Feedback can be delivered at any time in SemVP, and several key moments of delivery can be considered: delivering feedback at the student's request, after a fixed number of observations, after an inappropriate observation, etc..The educational impact of these variations could reveal important findings in the field of e-learning and medical education.

The emotional and psychological aspects of virtual patients can also be a fruitful direction for research in medical education. SemVP's model could be extended to include emotional aspects of the patient's condition and responses. The resulting model would provide useful insight in the field of psychology education.

Finally, using SemVP as a teaching tool during the course of a semester in a medical school and measuring the qualitative and quantitative impact the system has on student's clinical reasoning abilities is likely to yield useful results for the advancement of the field of virtual patients and medical education.

Appendix A

Details of Ms. Matibunda's Case Data

Observation	Associated Symptom(s)	Associated Risk Factor(s)	Associated condition(s)
Questions			
1.Any stiffness in your hands?	Stiffness	–	Arthritis, complex regional pain syndrome
2.Any swelling in your hands?	Swelling	–	Arthritis, osteoarthritis
3.You mentioned that you are worried. Can you tell me what worries you specifically?	–	Family history of diabetes	Diabetes
4.Can you describe the sensation in your hand a bit more?	Paresthesia	–	Diabetes, MS, carpal tunnel syndrome, osteoarthritis, peripheral neuropathy, motor neuropathy, sensory neuropathy
5.Do you feel the pain on all fingers?	Pain in the median nerve	–	Carpal tunnel syndrome
6.Are your feet painful?	Pain in feet	–	Diabetes, peripheral neuropathy

7.Do you feel like your hands are getting weaker?	Muscle atrophy, fatigue	–	Complex regional pain syndrome, amyotrophic lateral sclerosis, peripheral motor neuropathy, kidney cancer, diabetes, anemia
8.Do you take any particular medication, vitamins, etc. ?	–	–	Excessive vitamin A consumption
9.How long have you had this problem?	Slow onset paresthesia	–	Diabetes
10.Are you feeling more thirsty these days?	Excessive thirst	–	Diabetes
11.Have you been inexplicably loosing any weight recently?	Weight Loss	–	Kidney Cancer, arthritis, type 2 diabetes, amyotrophic lateral sclerosis, pancreatic cancer
12.How much alcohol do you drink?	–	Alcohol abuse	Kidney cancer, renal failure, and diabetes

13.Do you smoke?	–	Smoking	Osteoporosis, Heart disease, kidney cancer, lung cancer, angina, pancreatic cancer, hypertension
14.Is it painful on both hands?	Bilateral hand pain	–	Repetitive strain injury, diabetes type 1 and 2, complex regional pain syndrome, arthritis, osteoarthritis, peripheral neuropathy
15.What do you eat?	–	–	A healthy diet may prevent diabetes, coronary heart disease, anemia, obesity, hypertension, and osteoporosis.
16.Is you vision blurry?	Blurry vision	–	Type 2 diabetes, prediabetes, and brain tumor.
17.Does any of your relatives suffer from MS?	–	Family history of MS	MS
18.Does any of your relatives suffer from diabetes?	–	Family history of diabetes	Diabetes

19.Any back pain?	Back pain	–	Bladder cancer, Kidney cancer, osteoporosis, nerve entrapment
20.Are you feeling more hungry these days?	Excessive hunger(polyphagia)	–	
21.Do you do a lot of manual work (sewing, painting, etc.) at work or at home?	–	Repetitive hand movements	Carpal tunnel syndrome, complex regional pain syndrome
22.Any headache?	Headache	–	Influenza, brain tumor, anemia, hypertension, coronary heart disease, diabetes
23.Any shortness of breath, cough or chest pain?	Chest pain, shortness of breath	–	Angina, coronary heart disease
24.Are your legs painful?	Leg pain	–	Diabetes, peripheral neuropathy
25.Do you exercise?	–	–	May prevent diabetes type 1 and 2, prediabetes, and coronary heart disease
26.Do you sometimes forget things more than you used to?	Memory loss	–	MS, anemia

Examinations			
27.Examine Hands	swelling, stiffness, rash	–	Arthritis, osteoarthritis, complex regional pain syndrome
28.Capillary refill	Increased capillary refill time	–	Coronary heart disease, Raynaud's syndrome
29.Measure Height	–	Obesity	Diabetes, arthritis
30.Measure Weight	–	Obesity	Diabetes, arthritis
31.Measure blood pressure / pulse	–	–	Hypertension
Lab Tests			
32.HbA1c test	High HbA1c level	–	Diabetes
33.Random Plasma glucose	Increased random blood glucose level	–	Diabetes
34.Urine Ketone	Increased ketone levels	–	Diabetes
35.Creatinine	High creatinine level	–	Diabetes, renal failure
36.Fasting blood sugar	Increased fasting blood glucose level	–	Diabetes
37.Triglycerides	High triglyceride level	–	Diabetes
38.Postprandial Glucose Test	Increased blood glucose level	–	Diabetes

39.Urine Glucose	Glucose in urine	–	Diabetes, renal failure, kidney cancer
40.Heamoglobin	Decreased Mean Cell Haemoglobin	–	Anemia
41.Insuline level	Low insulin level	–	Diabetes, pancreatic cancer
42.LDL Cholesterol	–	Increased LDL cholesterol	Cardiovascular disease, diabetes

Table A.1: Data representing Ms. Matibunda

Appendix B

Additional Result (Final Study)

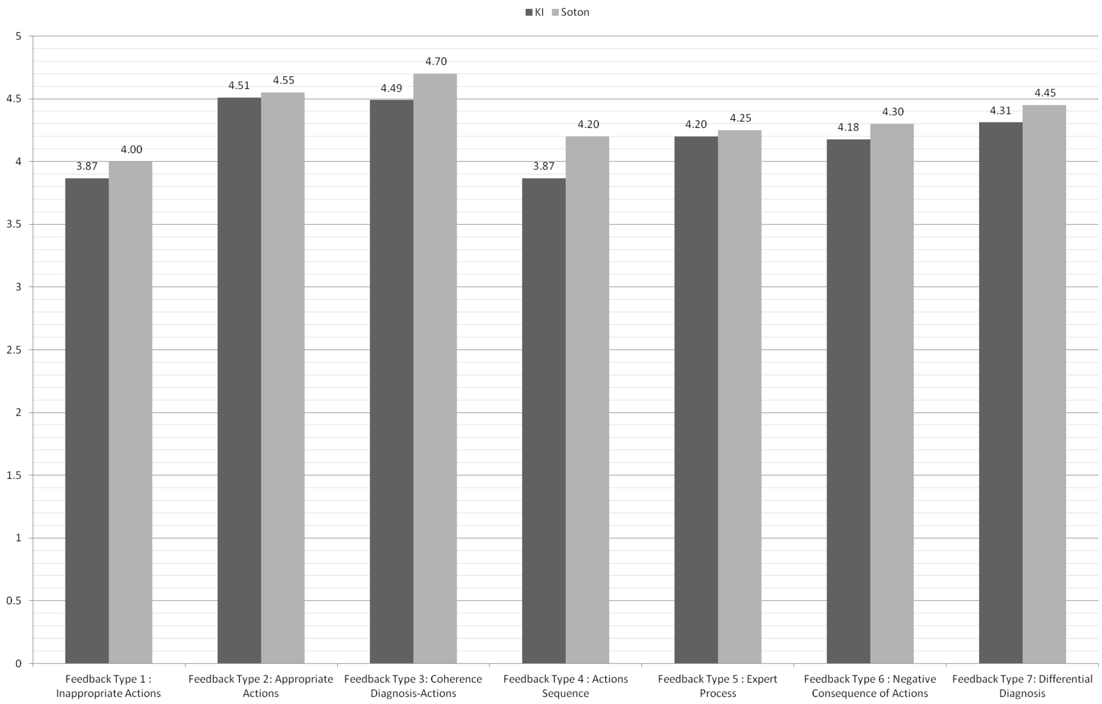


Figure B.1: Mean ratings of perceived usefulness for each feedback type by school, before using SemVP

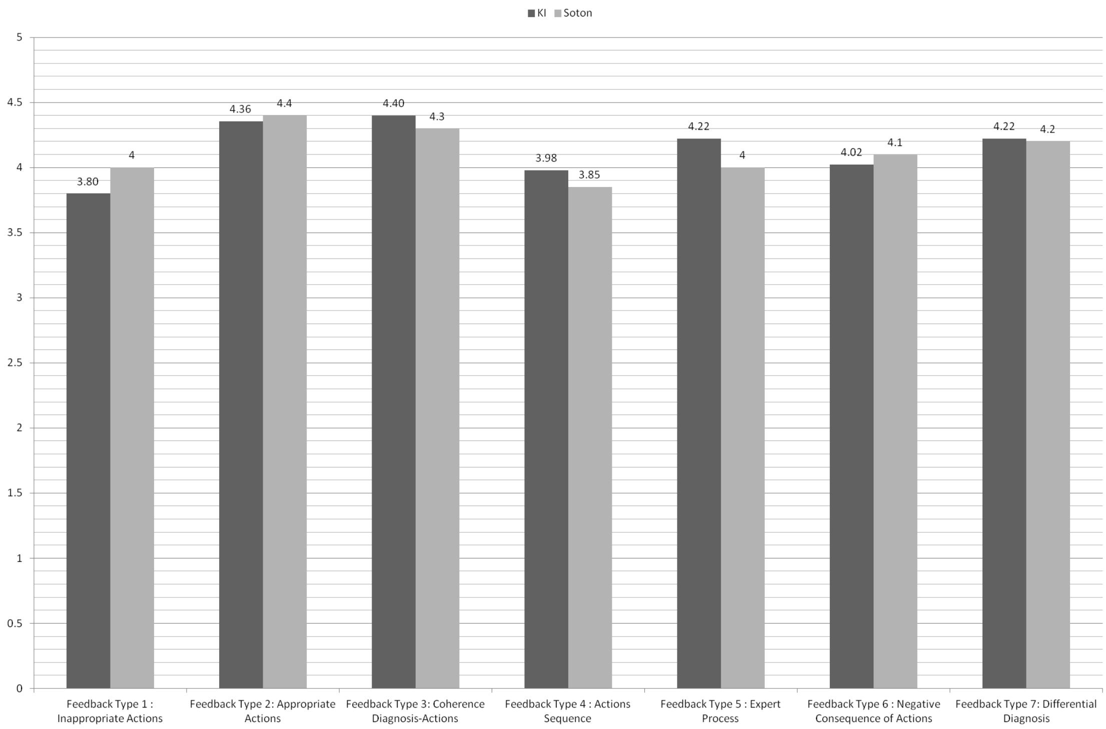


Figure B.2: Mean ratings of perceived usefulness for each feedback type by school, after using SemVP

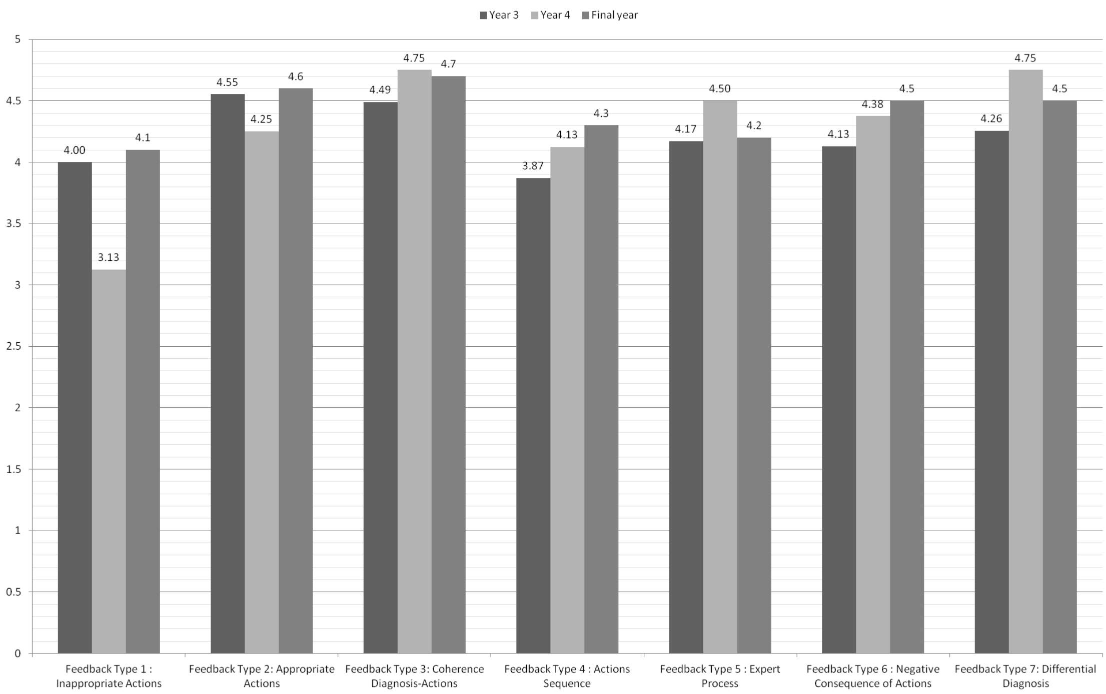


Figure B.3: Mean ratings of perceived usefulness for each feedback type by year group, before using SemVP

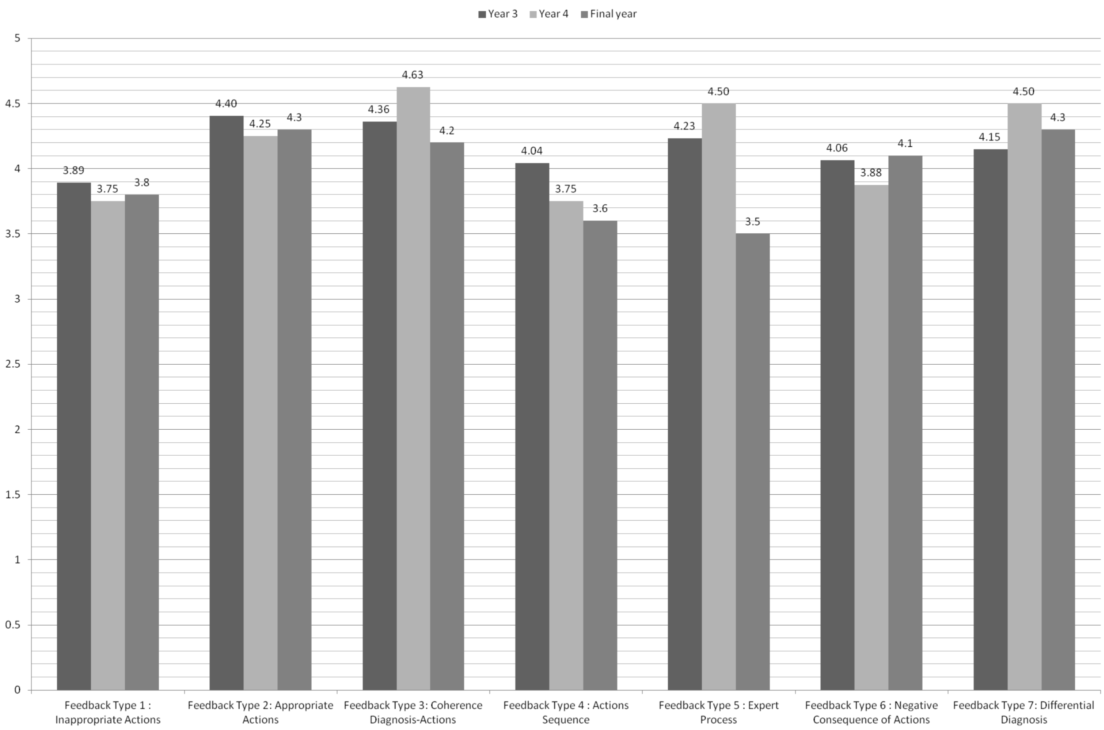


Figure B.4: Mean ratings of perceived usefulness for each feedback type by year group, after using SemVP

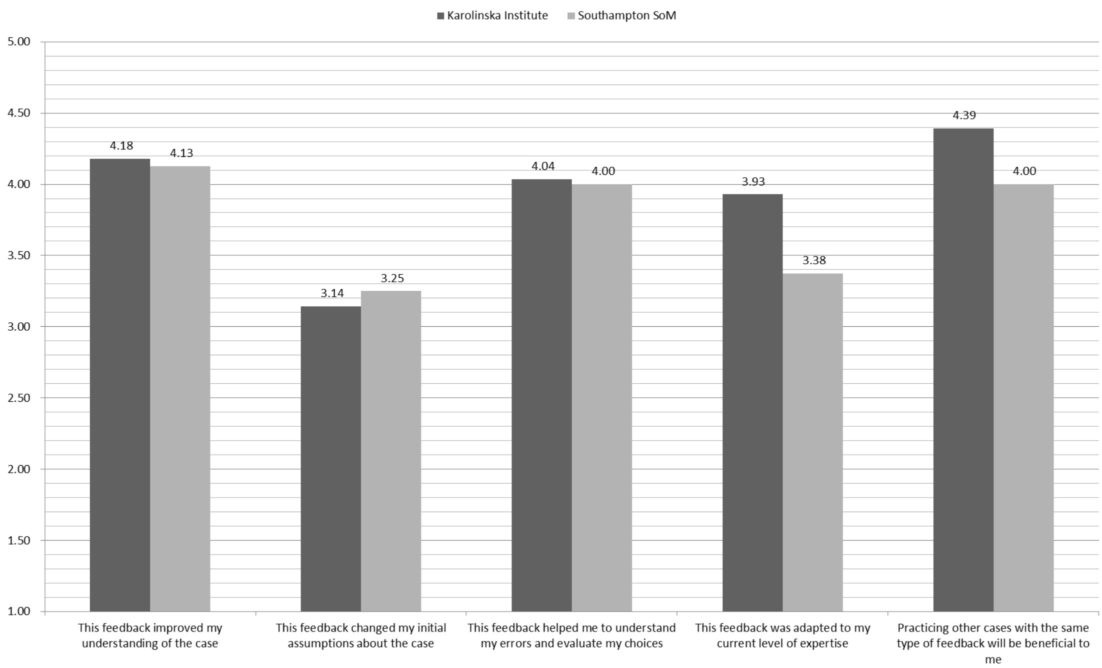


Figure B.5: Mean ratings by school for dynamic feedback

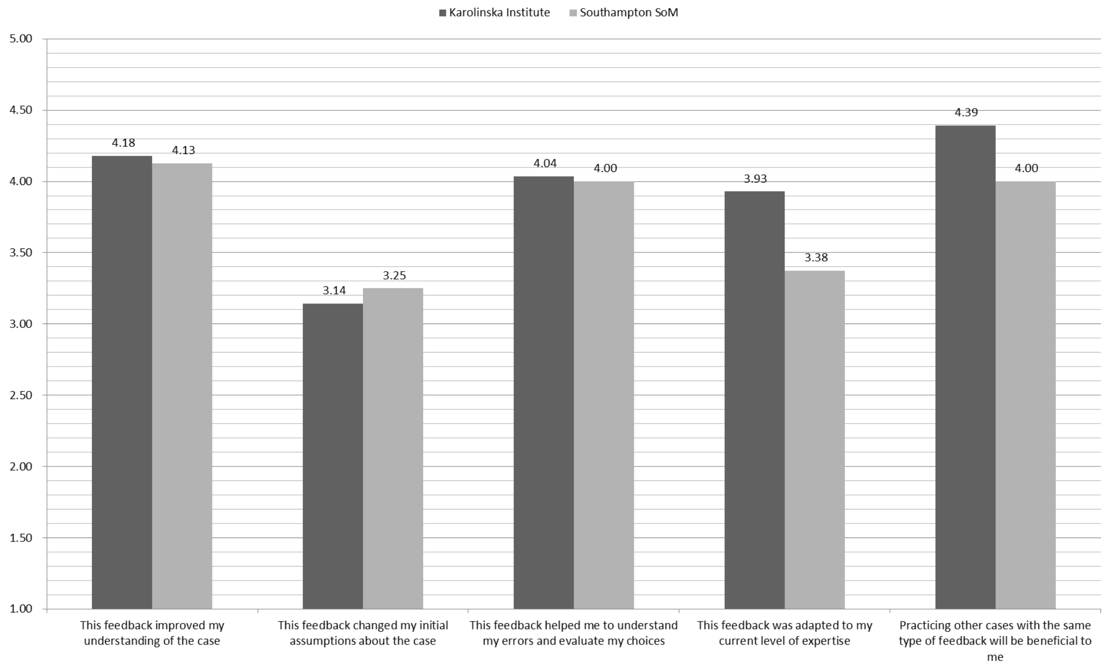


Figure B.6: Mean ratings by school for static feedback

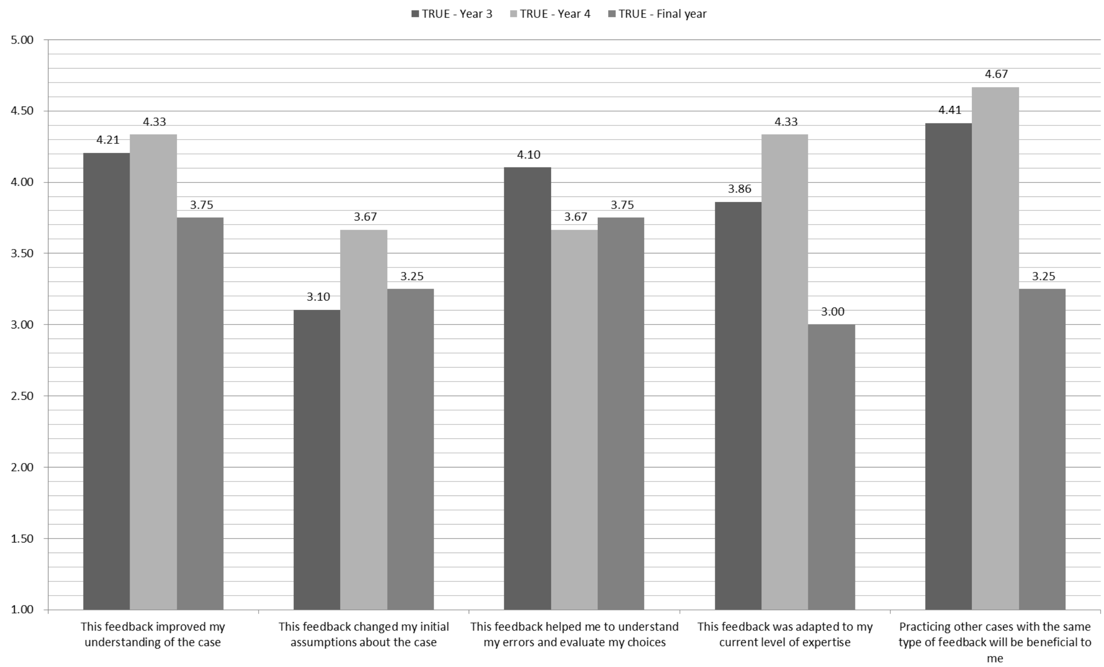


Figure B.7: Mean ratings by year group for dynamic feedback

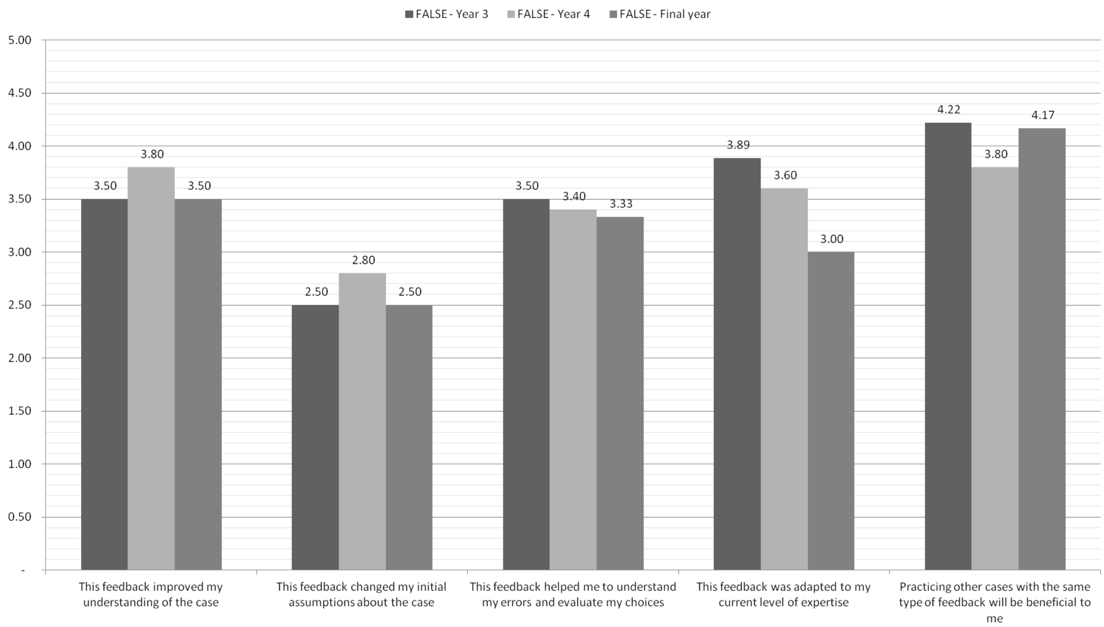


Figure B.8: Mean ratings by year group for static feedback

References

M. Albanese. Problem-based learning. In Tim Swanwick, editor, *Understanding Medical Education*, pages 37–52. Wiley-Blackwell, 2011.

G. Antoniou and F. Van Harmelen. *A Semantic Web primer - 2nd ed.* MIT Press, 2008.

M. Begg, R. Ellaway, D. Dewhurst, and H. MacLeod. Transforming professional healthcare narratives into structured game-informed learning activities. *Innovate Online*, 3: 7, August 2007.

N. Benedict. Virtual patients and problem-based learning in advanced therapeutics. *American Journal of Pharmaceutical Education*, 74(8):143, October 2010.

T. Berners-Lee. *Weaving the Web*. Harper Collins, 1999.

T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.

P.V. Biron and A. Malhotra. XML Schema Part 2: Datatypes Second Edition. http://www.w3.org/TR/xmlschema-2/. *W3C Recommendation*, 2004.

P. Black and D. Wiliam. *Assessment and Classroom Learning*. Kluwer Academic Publishers, 2006.

H.P. Boshuizen, H.G. Schmidt, E.J.F.M. Custers, and M.W. Van De Wiel. Knowledge development and restructuring in the domain of medicine: The role of theory and practice. *Learning and Instruction*, 5(4):269–289, 1995.

M. Botezatu, H.K. Hult, and U.G. Fors. Virtual patient simulation: what do students make of it? A focus group study. *BMC Medical Education*, 10(1):91, December 2010.

T. Bray, J. Paoli, C. Sperberg-McQueen, E. Maler, and F. Yergeau. Extensible Markup Language (XML) 1.0 (Fifth Edition). http://www.w3.org/TR/REC-xml/. November 2008.

D. Brickley and R.V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. http://www.w3.org/TR/rdf-schema/. *W3C Recommendation*, February 2004.

S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

P.M. Casey, A.R. Goepfert, E.L. Espey, M.M. Hammoud, J.M. Kaczmarczyk, N.T. Katz, J.J. Neutens, F.S. Nuthalapaty, and E. Peskin. To the point: reviews in medical education - the Objective Structured Clinical Examination. *American Journal of Obstetrics and Gynecology*, 200(1):25–34, January 2009.

M. Cavazza and A. Simo. A virtual patient based on qualitative simulation. In *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*, pages 19–25, New York, USA, January 2003. ACM.

B. Charlin, H.P. a Boshuizen, E.J. Custers, and P.J. Feltovich. Scripts and clinical reasoning. *Medical Education*, 41(12):1178–1784, December 2007.

S. Choi, A. Webb, E. Ault, and J.R. Duboc. Virtual Patients: Situated Learning For Pre-clinical Year 1 Students. In *Proceedings of the 2nd International Conference on Virtual Patients and MedBiquitous Annual Conference*, London, United Kingdom, 2010.

K.G. Clark, L. Feigenbaum, and E. Torres. SPARQL Protocol for RDF. http://www.w3.org/TR/rdf-sparql-protocol/. *W3C Recommendation*, January 2008.

S. Coderre, H. Mandin, P.H. Harasym, and G.H. Fick. Diagnostic reasoning strategies and diagnostic success. *Medical Education*, 37(8):695–703, August 2003.

D.A. Cook and D.M. Dupras. A practical guide to developing effective web-based learning. *Journal of General Internal Medicine*, 19(6):698–707, June 2004.

D.A. Cook and M.M. Triola. Virtual patients: a critical literature review and proposed next steps. *Medical Education*, 43(4):303–311, April 2009.

R. Ellaway. OpenLabyrinth: An abstract pathway-based serious game engine for professional education. In *Fifth International Conference on Digital Information Management (ICDIM)*, pages 490–495, Thunder Bay, Canada, July 2010.

R. Ellaway, C. Candler, P. Greene, and V. Smothers. *An Architectural Model for MedBiquitous Virtual Patients*. MedBiquitous Consortium, October 2006.

R. Ellaway, V. Smothers, and B. Azan. *MedBiquitous Virtual Patient Specifications and Description Document*. MedBiquitous Consortium, April 2010.

A.S. Elstein, A. Schwartz, and A. Schwarz. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal (Clinical research edition)*, 324(7339):729–732, March 2002.

R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol HTTP/1.1. http://www.rfc-editor.org/rfc/rfc2616.txt. *IETF Request For Comments 2616*, June 1999.

M.R. Fischer. CASUS - An authoring and learning tool supporting diagnostic reasoning. *Zeitschrift für Hochschuldidaktik*, 1:87–98, 2000.

M.H. Friedman, K.J. Connell, A.J. Olthoff, J.M. Sinacore, and G. Bordage. Medical student errors in making a diagnosis. *Academic Medicine*, 73(10):19–21, October 1998.

E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design patterns : elements of reusable object-oriented software*. Addison Wesley, 1995.

A. Gangemi. Ontology Design Patterns for Semantic Web Content. In *Proceedings of the Fourth International Semantic Web Conference*, Lecture Notes in Computer Science, pages 262–276. Springer Berlin / Heidelberg, 2005.

M. Gartmeier, J. Bauer, H. Gruber, and H. Heid. Negative knowledge: understanding professional learning and expertise. *Vocations and Learning: Studies in Vocational and Professional Education*, 1(30):87–103, February 2008.

G. Gauthier, L. Naismith, S.P. Lajoie, and J. Wiseman. Using Expert decision maps to promote reflection and self-assessment in medical case-based Instruction. In *9th International Conference on Intelligent Tutoring Systems*, 2008.

D.H. Gil, M. Heins, and P.B. Jones. Perceptions of medical school faculty members and students on clinical clerkship feedback. *Journal of Medical Education*, 59(11 Pt 1): 856–864, November 1984.

C. Golbreich, S. Zhang, and O. Bodenreider. The foundational model of anatomy in OWL: Experience and perspectives. *Semantic Web for Life Sciences*, 4(3):181–195, 2006.

M.E. Gozum. Emulating cognitive diagnostic skills without clinical experience: a report of medical students using Quick Medical Reference and Iliad in the diagnosis of difficult clinical cases. In *Annual Symposium on Computer Application in Medical Care*, 1994.

P. Grenon, B. Smith, and L. Goldberg. Biodynamic ontology: applying BFO in the biomedical domain. *Studies in Health Technology and Informatics*, 102:20–38, 2004.

I. Hege, V. Ropp, M. Adler, K. Radon, G. Mäsch, H. Lyon, and M.R. Fischer. Experiences with different integration strategies of case-based e-learning. *Medical Teacher*, 29(8):791–797, October 2007.

S. Huwendiek, B.A. De Leng, N. Zary, M.R. Fischer, J.G. Ruiz, and R. Ellaway. Towards a typology of virtual patients. *Medical Teacher*, 31(8):743–748, August 2009.

J.P. Kassirer and R.I. Kopelman. Cognitive errors in diagnosis: instantiation, classification, and consequences. *American Journal of Medicine*, 86(4):433–441, April 1989.

P. Kenny, T.D. Parsons, J. Gratch, A. Leuski, and A.A. Rizzo. Virtual patients for clinical therapist skills training. In *IVA 07: 7th International Conference on Intelligent Virtual Agents (LNCS 4722)*, Paris, France, 2007. Springer-Verlag.

A. Kononowicz, J. Heid, J. Donkers, I. Hege, L. Woodham, and N. Zary. Development and validation of strategies to test for interoperability of virtual patients. *Studies in Health Technology and Informatics*, 150:185–189, 2009.

P. Langley, H.A. Simon, G.L. Bradshaw, and J.M. Zytkow. *Scientific discovery: computational explorations of the creative processes.* MIT Press, 1987.

D.P. Larsen, A.C. Butler, and H.L. Roediger. Test-enhanced learning in medical education. *Medical Education*, 42(10):959–966, October 2008.

A. Leff and J.T. Rayfield. Web-application development using the Model/View/Controller design pattern. In *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference*, pages 118–127. IEEE Comput. Soc, 2001.

H.C. Lyon and M.R. Fisher. The Genesis of case-based authoring shell: from PlanAlyzer to VideoAtlas to CIMAS to CASUS/ProMediWeb and lessons learned from each. In *Slice of Life 2001 and Computers in Healthcare Education Symposium*, pages 195–206, Medical Faculty, Ludwig-Maximilian-University, Munich, F.R. Germany, August 2001.

H.C. Lyon, J.C. Healy, J.R. Bell, J.F. O'Donnell, E.K. Shultz, M. Moore-West, R.S. Wigton, F. Hirai, and J.R. Beck. PlanAlyzer, an interactive computer-assisted program to teach clinical problem solving in diagnosing anemia and coronary artery disease. *Academic Medicine*, 67(12):821–828, December 1992.

K. Malterud. Qualitative research: standards, challenges, and guidelines. *Lancet*, 358 (9280):483–488, August 2001.

H. Mandin, A. Jones, W. Woloschuk, and P. Harasym. Helping students learn to think like experts when solving clinical problems. *Academic Medicine*, 72(3):173–179, March 1997.

F. Manola and E. Miller. RDF primer. http://www.w3.org/TR/REC-rdf-syntax/. *W3C Recommendations*, February 2004.

S.P. Marshall. *Schemas in problem solving.* Cambridge University Press, 1995.

C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. Wonderweb eu project deliverable D18: The wonderweb library of foundational ontologies. *Trento, Italy: Laboratory For Applied Ontology-ISTC-CNR*, 2003.

C. Matuszek, J. Cabral, M. Witbrock, and J. Deoliveira. An introduction to the syntax and content of Cyc. In *Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*. AAAI Press, March 2006.

D.L. McGuinness and F. van Harmelen. OWL overview. http://www.w3.org/TR/owl-features/. *W3C Recommendation*, February 2004.

J. McIlwrick, B. Nair, and G. Montgomery. "How am I doing?": many problems but few solutions related to feedback delivery in undergraduate psychiatry education. *Academic Psychiatry, the Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, 30(2): 130–135, March 2006.

J. McKendree. e-learning. In Tim Swanwick, editor, *Understanding Medical Education*, chapter 11, pages 151–163. Wiley-Blackwell, 2011.

J.J.G. Merriënboer, R.E. Clark, and M.B.M. de Croock. Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research & Development*, 50(2):39–61, 2002.

E.C. Molina. Bayes' theorem: an expository presentation. *The Annals of Mathematical Statistics*, 2(1):23–37, 1931.

R. Möller and V. Haarslev. Tableau-Based Reasoning. In *Handbook on Ontologies*, pages 509–528. Springer, 2009.

D.R. Needham and I.M. Begg. Problem-oriented training promotes spontaneous analogical transfer: memory-oriented training promotes memory for training. *Memory & cognition*, 19(6):543–557, November 1991.

G.R. Norman. Research in clinical reasoning: past history and current trends. *Medical Education*, 39(4):418–427, April 2005.

G.R. Norman. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education Theory and Practice*, 15(5):625–632, December 2010.

G.R. Norman and H.G. Schmidt. The psychological basis of problem-based learning: a review of the evidence. *Academic Medicine*, 67(9):557–565, September 1992.

G. Pólya. *How To Solve It: A New Aspect of Mathematical Method*. Princeton University Press, 1957.

T. Poulton and C. Balasubramaniam. Virtual patients: a year of change. *Medical Teacher*, 33(11):933–937, January 2011.

T. Poulton, E. Conradia, S. Kavia, J. Rounda, and S. Hilton. The replacement of "paper" cases by interactive online virtual patients in problem-based learning. *Medical Teacher*, 31(8):752–758, August 2009.

E. Prud'hommeaux and A. Seaborne. SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/. *W3C Recommendation*, January 2008.

X.A. Qu, R.C. Gudivada, A.G. Jegga, E.K. Neumann, and B.J. Aronow. Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinformatics*, 10(Suppl), May 2009.

A.L. Rector, J.E. Rogers, P.E. Zanstra, and E. Van Der Haring. OpenGALEN: open source medical terminology and tools. In *Proceedings of the AMIA Symposium*, page 982, 2003.

R. Rikers, H. Schmidt, and H. Boshuizen. Knowledge Encapsulation and the Intermediate Effect. *Contemporary Educational Psychology*, 25(2):150–166, April 2000.

C. Rosse and J.L.V. Mejino. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36(6):478–500, December 2003.

F. Ruderich, M. Bauch, M. Haag, J. Heid, F. J. Leven, R. Singer, H.K. Geiss, J. Jünger, and B. Tönshoff. CAMPUS - a flexible, interactive system for web-based, problem-based learning in health care. *Studies in Health Technology and Informatics*, 107(Pt 2):921–925, February 2004.

H. Saarinen-Rahiika and J.M. Binkley. Problem-based learning in physical therapy: a review of the literature and overview of the McMaster University experience. *Physical Therapy*, 78(2):195–211, February 1998.

M. Sawdon and F. Curtis. Building interactive clinical case studies using and evaluating online software. In *Association for Medical Education in Europe (AMEE) International Meeting*, Glasgow, UK, 2010.

H.G. Schmidt. Problem-based learning: rationale and description. *Medical Education*, 17(1):11–16, January 1983.

H.G. Schmidt and H.P. Boshuizen. On acquiring expertise in medicine. *Educational Psychology Review*, 5(3):205–221, 1993.

H.G. Schmidt, G.R. Norman, and H.P. Boshuizen. A cognitive perspective on medical expertise: theory and implication. *Academic Medicine*, 65(10):611–621, October 1990.

A. Sender Liberman, M. Liberman, Y. Steinert, P. McLeod, and S. Meterissian. Surgery residents and attending surgeons have different perceptions of feedback. *Medical Teacher*, 27(5):470–472, August 2005.

N.R. Shadbolt. Eliciting expertise. In *Evaluation of Human Work*. Taylor & Francis Ltd, 2005.

D.A. Sloan, M.B. Donnelly, R.W. Schwartz, J.L. Felts, A.V. Blue, and W.E. Strodel. The Use of the Objective Structured Clinical Examination (OSCE) for Evaluation and Instruction in Graduate Medical Education. *Journal of Surgical Research*, 63(1): 225–230, June 1996.

B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.A. Sansone, R.H. Scheuermann, N. Shah, R.L. Whetzel, and S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, November 2007.

C. Snyder. *Paper Prototyping*. Morgan Kaufmann, 2003.

M. Toro-Troconis, U. Mellström, M. Partridge, and M. Barrett. An architectural model for the design of game-based learning activities for virtual patients in Second Life. In *European Conference on Game-Based Learning*, The Universitat Oberta de Catalunya (UOC), Barcelona, October 2008.

J.M.M. Van de Ridder, K.M. Stokking, W.C. McGaghie, and O.T.J. Ten Cate. What is feedback in clinical education? *Medical Education*, 42(2):189–197, January 2008.

T. Van Gog, F. Paas, and J.J.G. van Merriënboer. Process-oriented worked examples: improving transfer performance through enhanced understanding. *Instructional Science*, 32(1-2):83–98, January 2004.

D.T. Vernon and R.L. Blake. Does problem-based learning work? A meta- analysis of evaluative research. *Academic Medicine*, 68(7):550–563, July 1993.

W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). http://www.w3.org/TR/owl2-overview/. *W3C Recommendation*, December 2012.

D.F. Wood. Formative assessment. In Tim Swanwick, editor, *Understanding Medical Education*, pages 259–270. Wiley-Blackwell, 2011.

N. Zary, G. Johnson, J. Boberg, and U. Fors. Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment - Web-SP. *BMC Medical Education*, 6(1), February 2006.