

Chapter 5 Formative Evaluation

The purpose of this thesis has been to explore how reflection can be supported, while medical students or postgraduate medical practitioners undertake a series of simulated consultations, involving the diagnosis and management of chronic illness. Consistent with the approach of Alexander, McKenzie, and Geissinger (1998), formative evaluation was undertaken in three cycles at the design and development stages. The first two cycles were undertaken with two main aims. The first was to uncover and correct problems with the user interface. This was done to ensure SIMPRAC was sufficiently usable, before undertaking a study involving the intended end users. The second aim was to obtain feedback on the review component that was designed to support reflection, seek suggestions for improvement, as well as have an opportunity to see it in use, as a basis for our own reflections on its effectiveness, and ways to improve this critical aspect of SIMPRAC. The third cycle was much larger, and included qualitative and quantitative data involving the intended end users, and assessed SIMPRAC against its design goals. This chapter presents the results of the first two evaluation cycles, while the following chapter reports on the last evaluation.

5.1 Design

The design for SIMPRAC was founded on the model of the consultative process illustrated in Figure 1 on page 15. The patient interaction components of the software were designed to support the consultation process, while an additional set of tools and approaches were developed to support reflection.

Chapter 3 describes the user interface to the various elements of the software. Obviously, much of this interface is devoted to the data gathering and management

selection elements of the model. Thus the history taking, physical examinations, and investigation components are all related to the process of data gathering, while the management selection screen (Figure 20, page 57) has been developed to support patient management by the user. The various hypotheses generated by the user are captured using the hypothesis screen (Figure 22, page 61).

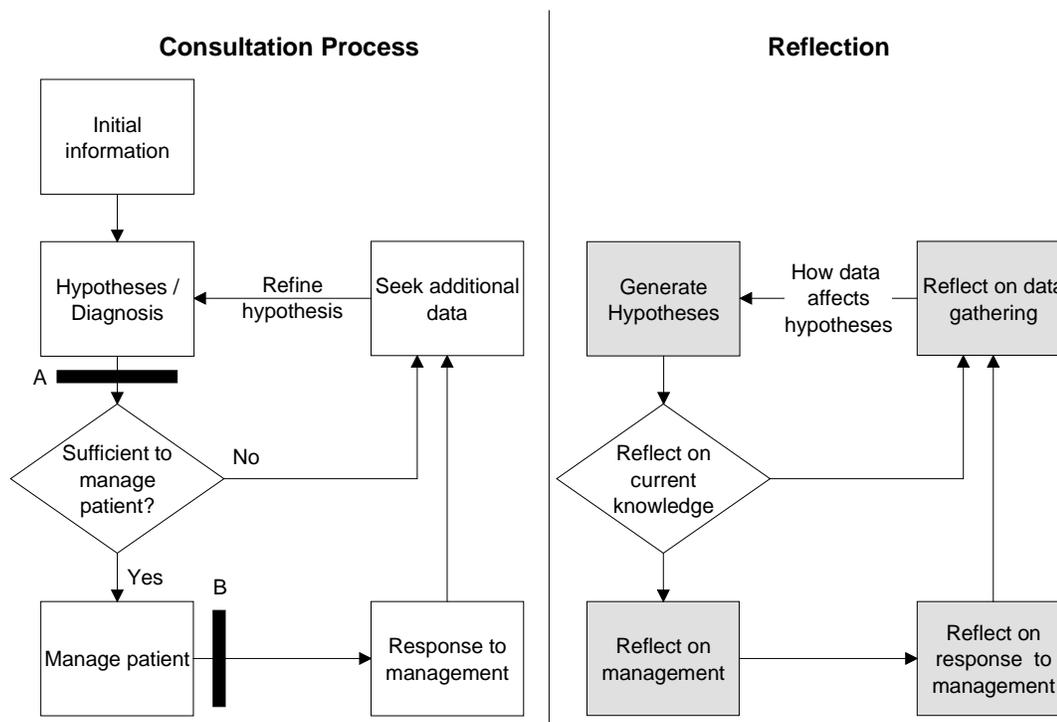


Figure 37: Model of the consultation process. Steps A and B indicate where the consultation process is interrupted to promote reflection.

With the initial design, the response to management can be discovered by the user in two ways. First, the user might be informed of any changes to the patient's condition at the beginning of the following consultation. However, in the first case, this was reserved for major events that the author wanted to bring to the user's attention. The second means, by which the user can learn of the patient's response, is through appropriate data gathering. This might involve a change in the patient's symptoms (history taking), signs

(physical examination), or the results of investigations. For example, use of an appropriate therapeutic agent for lowering serum triglycerides, results in a reduction in the serum triglyceride concentration on retesting.

Support for reflection was considered in several ways. Firstly, as stated above, generation of a set of hypotheses is a reflective process in itself. The question then becomes, when and how these should be updated and refined. When interacting with a patient, this is an ongoing process, although it may not be a specific conscious activity. Each activity aimed at acquiring additional information may potentially affect the hypotheses held by the clinician. While one could contemplate making the user tie each question, examination, or investigation to their hypotheses, and state how each of these supported or refuted the hypothesis in question, this would be a very tedious process, unlikely to be followed by most users. Based on the experience of the author, most consultations, outside of the acute emergency arena, are undertaken in a staged process. Initially, a history is taken from the patient before a physical examination is performed. The physical examination is generally followed by requests for various investigations, but may also lead to further history-taking. Similarly, the results of investigations may lead to further history-taking and physical examination before a management strategy is decided upon. During an acute presentation, each of these processes may be undertaken concurrently. With this staged approach in mind, and as a compromise to constant interruption, the system was developed so that users were prompted to update their hypotheses before they could perform a physical examination, order an investigation, or select a management option for the first time in each consultation. This intervention is marked 'A' in Figure 37.

Having developed a set of hypotheses, one can then reflect on the decision as to whether there is sufficient information to manage the patient or institute a trial of management. Again, it was thought that asking the user to do this as new information was acquired would be too intrusive. Furthermore, given the fact that the user was already being prompted to review their hypotheses, it was felt that an explicit process or procedure to encourage the user to ask this question and reflect was not warranted. Nevertheless, implicit support, in the form of the user notes and medical record components, was provided. It was envisaged that the user notes component would enable the user to list any learning goals they might have, while the medical record component, in addition to being an *aide memoir*, would enable the user to record the patient management issues needing to be addressed.

With the initial design, a single component was included in SIMPRAC to encourage users to reflect on the appropriateness of their data gathering and management selections. This involved a graphical summary of the proportion of critical, relevant or irrelevant actions selected by the users, with the ability to compare their pattern of activity with the average for their peers, or an expert in that particular knowledge domain (Figure 26 on page 66, Figure 27, and Figure 28). As the patient's response to therapy was not only given before commencing another consultation, but also elicited during the following consultation, this feedback and opportunity for reflection was made available after the management options for the current consultation had been selected, and before the following consultation. This intervention is marked 'B' in Figure 37.

5.2 Methods

All three cycles used a simplified think-aloud protocol as described by Nielsen (Nielsen, 1993). During the think-aloud sessions, each participant was asked to think-aloud, stating what they were doing and thinking as they used the interface. The experimenter merely observed and recorded the activity and, in cases where the participants appeared to spend time thinking but not externalising, encouraged them to express what they were thinking. If the user appeared to become stuck at some point, the experimenter helped them to move on, recording this fact in the detailed record of the session. This method enables the interface designer to obtain qualitative data on how the application is perceived by the user, and whether the application is used in a manner intended by the design team. In contrast to more traditional think-aloud experiments, only a written record was kept and the experiment did not record the interaction on audio or video tape. Example think-aloud records can be found at Appendix F on page 249 and Appendix G on page 254. A questionnaire was prepared for the third development cycle and is reproduced at Appendix C on page 241.

Evaluation Cycle One

The aim of the first evaluation cycle was to assess the usability of the application and inform further development. This involved use of SIMPRAC by two medical graduates. The first user was this author's supervisor, and is denoted as User A in Table 4. This individual was not in clinical practice but had extensive input into the design of the application. User A was chosen because he had a good understanding of the goals of the application and its intended use. The second user, User B in Table 4, was a medical graduate with ten years postgraduate experience, who was in their final year of advanced training in chemical pathology. This user was chosen because she was experienced in the management of lipid disorders, was able to give feedback on the

medical aspects of the case, but was unfamiliar with the design SIMPRAC. The simplified think-aloud evaluations were performed on separate days at separate locations using different computers.

Table 4: Client software configuration for each user.

User	Status	Cycle	Operating System	Web Browser	Java Runtime Environment
A	Supervisor	1	Redhat Linux 7.3	Mozilla 1.1	1.4.0
B	Advanced Trainee One	1	Windows 98	Internet Explorer 5.5	1.3
C	Advanced Trainee One	2	Windows 98	Internet Explorer 5.5	1.3
D	Advanced Trainee Two	2	Windows XP	Internet Explorer 6.0	1.4.1

Before commencing the evaluation, each user was informed that the purpose of the evaluation was to assess the ease of use and structure of the application, as well as the structure of the first case. The user was specifically informed that they were not being evaluated, and that the information was only being used to improve the case simulation. The various client configurations for each of the computers used by the users are summarized in Table 4. For both the first and second evaluation cycles the server software was run on a remote computer under RedHat Linux 7.3. In contrast, during the third evaluation cycle, where users were located at multiple sites, often without convenient access to an appropriately configured computer, the server software was run locally on the author's notebook computer.

Evaluation Cycle Two

Evaluation Cycle Two involved two advanced chemical pathology trainees, both with at least 10 years postgraduate medical experience, and both considered experts in the management of hyperlipidaemia. The first user in this evaluation was the same trainee who volunteered to be involved in the first evaluation. This second evaluation was

undertaken nine months after the first evaluation. Therefore this user did not recall in great depth, what actions had been undertaken in the previous exposure to SIMPRAC. Both users were chosen because they had experience in the management of these types of clinical problems, were able to give feedback on the medical aspects of the case, but were not overly familiar with the design of the application.

These two simplified think-aloud evaluations were both performed on the same day at the user's desk using different computers (see details in Table 4). Following the think-aloud evaluation, the users were asked to review the evaluation questionnaire. As with the first evaluation cycle, before starting the evaluation, each user was informed that the purpose of the evaluation was to assess the ease of use and structure of the application, as well as the structure of the first case. The user was specifically informed that they were not being evaluated, and that the information was only being recorded to improve the application.

Table 5: Summary Data for evaluations A and B. The duration is the time taken in minutes for each consultation. The time in parentheses is the total time taken for the evaluation and is greater than the sum of the times taken for all consultations due to periods of discussion with the experimenter. Stream represents one of several possible states for each consultation (as described on page 58). Questions, examinations, investigations, and management are the number of each these actions that were selected in each consultation. N/A: not applicable as the evaluation was terminated.

User	Consultation	1	2	3	4
A	Duration (85 min)	47	16	3	N/A
	Stream	1	1	3	N/A
	Questions	41	3	1	N/A
	Examinations	25	0	0	N/A
	Investigations	8	5	4	N/A
	Management	1	2	Ceased	N/A
B	Duration (64 min)	48	2	2	N/A
	Stream	1	1	4	N/A
	Questions	10	0	1	N/A
	Examinations	27	0	0	N/A
	Investigations	10	3	2	N/A
	Management	2	3	Ceased	N/A

5.3 Evaluation Cycle One

From Table 5 it can be seen that both users completed three consultations. In each case, the first consultation took considerably more time, 47 and 48 minutes for User A and User B respectively, than the second or third consultation. For both users, this was a direct result of the number of actions required to diagnose the problem. Nevertheless, there were considerable differences between User A and User B. User A asked 41 questions, performed 25 examinations, and requested 8 investigations in the first consultation. User B, with 10 questions, asked far fewer questions than User A. However, both users performed roughly the same number of examinations and investigations in this consultation. The management strategy chosen by both users was equivalent at the end of the first consultation with both users having a stream state of 1 in both the first and second consultations. At the end of the second consultation, User A managed the patient with a HMG CoA Reductase Inhibitor, with the patient moving to stream 3. In contrast, User B used a fibrate, resulting in more effective lowering of the triglyceride concentration, and moved to stream 4.

5.3.1 Evaluation A

The first evaluation began with the user referred to as 'A'. The evaluation was undertaken in a pathology research laboratory, on one of the laboratory's personal computers. The evaluation session took 85 minutes to complete. A more detailed coverage of the think-aloud session can be found at Appendix D on page 246.

User A managed to register with the application and select the appropriate case without much difficulty. However, he did indicate that the peer group / professional background selection was not obvious, when it was included on the same screen as the cases to be selected. This user suggested it be moved to a new screen on its own.

User A was able to interact with the patient appropriately but did make a number of suggestions that are listed below. All of these suggestions were included in the final version of the software used in the third evaluation cycle.

1. Separate the peer group selection from the professional background selection so that they appear on different screens.
2. If a question is asked from a question category list, such as family history, the list of questions for any given category should remain on screen after the question is answered, so that another question from the same category can be asked simply by clicking on another question hyperlink, rather than reselecting the category.
3. A save or submit button to be added to the hypothesis review screen, so that when the user enters their last hypothesis, the information is saved and the user redirected to the appropriate screen. Thus, if a user clicks the examination hyperlink before being redirected to the hypothesis review page, clicking on the save button redirects them to the patient examination screen.
4. The Java Applet used to play audio files should be modified to replace the existing three buttons; start, stop, and loop, with only two buttons. One to start the audio-clip in a loop, so that it plays repeatedly, and one to stop the sound.
5. Text interpretations of media objects should only be available after the user has submitted their own interpretation.
6. Instructions on how to select more than one test at a time from a list were adjacent to the list but in very small text. The size of the font should be increased.

7. On-screen instructions on how the management screen is to be used should be provided.
8. The review screen should open automatically at the end of a consultation.
9. Details of the current user were displayed on the first tab of the review screen. User A found this more confusing than useful, and thought it ought to be omitted.
10. Changing the comparison group or the stage of the consultation should automatically update the currently viewed graph.
11. While it was possible to see the proportion of critical, relevant, or irrelevant actions the user had selected in the original version, it was not possible to see any details. User A indicated that specific details of which actions were critical and relevant should be provided. He also stated that information should also be provided on items that were thought to be critical or relevant, but were not performed.
12. Medical Notes component and User Notes component were both accessible from a hyperlink in the left navigation frame labelled, "Notes". This link should be more descriptive.

While the think-aloud evaluation was being conducted, additional problems were observed by the author:

1. The number of management options available to the user was very limited.
2. When the same combination of chart type, comparison group, and stage of the consultation that had been previously charted, was again selected, a new chart was created rather than just displaying the old chart.
3. A hypertext link to the next consultation was not available from the end-of-consultation screen.

The last two items were simple interface issues and were corrected before Evaluation Cycle Two. The number of management items that could be selected was expanded in subsequent versions. However, the total number of options was still very small compared to the number of diagnostic actions available to the user. The other interface issues noted above were corrected before the third evaluation cycle.

5.3.2 Evaluation B

The second evaluation involved the user denoted as 'B' in Table 5. The evaluation was undertaken in user B's office, using their own computer. The evaluation session took 64 minutes to complete. This user was an advanced trainee in chemical pathology with experience in the management of hyperlipidaemia, at a tertiary referral level. While this user did not have experience in the management of Glycogen Storage Diseases, she did have experience in managing complex patients with elevated serum cholesterol and triglyceride concentrations.

Areas needing improvement, and user B's suggestions for this evaluation, are listed below.

1. Instructions relating to the use of the question and answer module of the history-taking screen should be provided at the top of the screen.
2. More specific questions relating to diet and exercise, with their appropriate responses, should be available to the user.
3. If an examination tool is selected and the body view is changed, for example from whole body to the abdomen, the examination tool should remain active, rather than having to reselect the tool again.

4. When searching for management options, the system should be more tolerant of minor misspellings. For example, the use of, “fibrate”, “fibrates”, or “fibrat” should all produce the same list of matching management options.
5. The review screen should open automatically at the end of a consultation.
6. At the end of consultation review, specific details on which diagnostic and management actions were critical and relevant should be provided. Information should also be provided on which critical or relevant actions were not chosen or performed.
7. When the review screen was closed a dialog box asking for confirmation was displayed. User B thought this should be removed as she was not sure whether answering yes to this would close just the review screen or the entire application.

While the think-aloud evaluation was being conducted, the author observed that the review applet, which had been compiled using version 1.4.1 of the java software development kit (Sun Microsystems), did not render correctly when displayed on user B's computer. User B had version 1.3 of the Java runtime environment installed. Based on this observation, the third evaluation cycle was undertaken using only version 1.4 or later of the Java Runtime Environment, or Java Software Development Kit (Sun Microsystems).

With the exception of items two and four, all the suggestions made by user B were implemented. For the purposes of the continuing software evaluation, it was thought that it was not necessary to add more specific questions relating to diet, as all the information available relating to diet and exercise could be included in the responses to the general questions. Furthermore, the addition of these questions to the database would have made authoring cases more challenging due to the increased variations.

With respect to the limitations of the free text search tool, the system was implemented using the third party Lucene library from the Apache Software Foundation. A decision was made to continue using this tool while working within its constraints, as the development of a more sophisticated tool was beyond the scope of this project.

In contrast to user A, user B was very much more focused on “solving” the case than providing feedback on the user interface. For example, they became frustrated with the inability to obtain a detailed dietary history from the patient. With the current system, the dietary history has been provided as a single response to the question, “Can you please describe your diet?”. In contrast, user B wanted to ask specifics about the amount of fat or oil ingested, and whether this fat was saturated, monounsaturated, or polyunsaturated. With the current level of sophistication it was not possible for them to do this. Similarly, user B wanted detailed information on the patient’s exercise pattern but this was not available.

The two users experienced some difficulty with the interface, although they differed in the type of difficulties experienced. For example, user A had some difficulty with the registration screen, and with selecting the management options they desired. In contrast, user B had some initial difficulty with registration and the history-taking screens. In both cases once the users had learnt how the interface was to be used, they had no difficulty using it, and few errors were observed. Presenting the case interaction according to common medical processes of history-taking, physical examination, investigation, and management, seemed to be intuitive. This is not surprising, given its simplicity, and the fact that the developers have a medical background. The fact that the

first user had no difficulty with the history interface being split into free text or category based selection was in part due to the fact that this user was responsible for making the suggestion that this be included as a design element. Thus the difficulty experienced by the second user may be more indicative of difficulties novice users might experience. In an attempt to resolve this issue, a detailed text description on how to use the history screen was included at the top of the screen.

From the perspective of our major concern, the support of reflection, both users were frustrated by the lack of feedback on what questions were considered critical but had not been asked. The second user also indicated that she would like feedback on what questions she had asked that were not considered relevant by the case author(s). Both of these observations suggest that the visual representation may be encouraging users to reflect. However, they also indicate that users would like additional specific information regarding an independent view of their actions. This specific information may form a perspective against which learners can base their reflections. As the review component was designed as the primary instrument for reflective learning, these facilities needed to be incorporated before further evaluation was undertaken. Gratifyingly, based upon observation of user B, it seemed that the “natural” feedback provided by changing patient outcomes, based on the management options chosen by the learner, was successful in directing the user to the desired treatment strategies.

Based on the feedback provided by these two users the following major changes were made:

1. The textual interpretations of various media objects such as heart sounds or X-rays were made available only after the user had submitted their own interpretation.

2. The review component was modified so that specific details were available to the user on the actions they had selected, and that the case authors considered critical, relevant, or not relevant to the diagnosis and management of the patient. On the same screen it was also possible to see those critical or relevant actions that were not performed.

In addition to the changes already mentioned, the review screen was modified to include an additional step, whereby users must categorise their actions in each of the history, examination, investigation, and management stages, as being critical, relevant, or not relevant in light of their current knowledge of the patient (see Figure 25 on page 65). This element was introduced to encourage the student to reflect on the usefulness and appropriateness of their actions in the completed consultation.

5.4 Development Evaluation Cycle Two

Table 6: Summary data for evaluations C and D. The duration is the time taken in minutes for each consultation. The time in parentheses is the total time taken for the evaluation and is greater than the sum of the times taken for all consultations due to periods of discussion with the experimenter. N/A: not applicable as the evaluation was terminated.

User	Consultation	1	2	3	4
C	Duration (60 min)	31	9	3	13
	Stream	1	4	4	6
	Questions	5	1	1	0
	Examinations	32	0	0	17
	Investigations	9	5	3	2
	Management	1	3	4	4
D	Duration (79 min)	69	N/A	N/A	N/A
	Stream	1	N/A	N/A	N/A
	Questions	30	Ceased	N/A	N/A
	Examinations	43	N/A	N/A	N/A
	Investigations	16	N/A	N/A	N/A
	Management	1	N/A	N/A	N/A

5.4.1 Evaluation C

This user was the same user as in evaluation B. However, this second evaluation was undertaken nine months after the first evaluation, and the user did not remember all the specific details of the case under study. Furthermore, the focus of this evaluation was to assess the user interface with the goal of making any necessary improvements before a larger user evaluation was undertaken. She was asked to register as a new user, with a different login name to that previously used. The evaluation was performed at the user's own desk, using her computer configured as outlined in Table 6. This think-aloud evaluation took 60 minutes to complete four simulated consultations. At the completion of the evaluation, the user was asked to review the questionnaire (see Appendix C on page 241) to be administered during the third evaluation cycle.

As shown in Table 6, User C completed all four consultations. While this user did not recall all the specific details of the case, there was evidence of a memory effect, as the first consultation was completed in much less time than in the first evaluation involving

this user, and effective lipid lowering medication was prescribed at the end of consultation one rather than at the end of consultation two. User C also asked fewer questions than she did at the previous encounter, performed slightly more examinations, and requested about the same number of investigations during the first consultation. As was the case for evaluations A and B, the second and subsequent consultations were much shorter, as a result of fewer diagnostic activities.

5.4.2 Evaluation D

This user was an advanced trainee in chemical pathology with significant experience in the management of hyperlipidaemia, at a tertiary referral level. The evaluation was performed at the registrar's own desk, using her own computer configured as outlined in Table 4 on page 90. This think-aloud evaluation took 79 minutes, and in contrast to the previous sessions, only a single consultation was completed. Subsequent consultations were not undertaken as the session had already taken much longer than the anticipated 60 minutes. At the completion of the evaluation, the user was asked to review the questionnaire to be administered during the third evaluation cycle. An example of the questionnaire can be found at Appendix C on page 241.

Summary of changes

The suggestions made by users C and D, and the action taken on these suggestions, are listed below.

1. The question and answer module needs to be improved to improve the accuracy with which it is able to find the appropriate question when entering free text. As is the case for the management search tool, the system was implemented using the third party Lucene library from the Apache Software Foundation. A decision was made to continue using this tool while working within its constraints, as the development of a more sophisticated tool was beyond the scope of this thesis.

2. Confusion was experienced by user D with the hypothesis page re-appearing when the examination, investigations, or management hyperlinks are clicked for the first time during each consultation. This was resolved following an explanation of the need to review hypotheses at the end of each diagnostic stage.
3. Users were uncertain how to get to the desired page once an hypothesis was added, as the user was not redirected automatically when the user had finished updating their hypotheses. A save button was added to the hypothesis review screen, so that when the user enters their last hypothesis, the information is saved and the user redirected to the appropriate screen. Thus, if a user clicks the examination hyperlink before being redirected to the hypothesis review page, clicking on the save button redirects them to the patient examination screen.
4. All tool and body-part combinations should have an appropriate response or finding. This was resolved by ensuring all possible combinations had a default value.
5. Not all investigation results were consistent with the case. For example, the low protein concentration seen on the liver function test was not consistent with the normal protein levels seen on the protein EPG. This arose because it had not been anticipated that an EPG would be requested. Discordant results were corrected as they became apparent.
6. User C did not use the review component at the end of the first, second or third consultations voluntarily. Instead, she proceeded to the next consultation. The software was modified to that the review screen opened automatically at the end of each consultation.
7. As previously noted in evaluation B (see page 96), the review applet, which was compiled under JSDK 1.4.1 did not display correctly on user C's computer

which was running JRE 1.3. Subsequent evaluations were conducted using JSDK 1.4 or later.

8. User C would have liked to investigate the outcome of alternate management strategies without having to login as a new user. This could not be changed without major architectural changes and was therefore not implemented. This issue is revisited and discussed in more detail in Chapter 7.
9. Having completed all four consultations, User C would also have liked a case summary indicating what the authors considered to be the “right” answer or answers. This was considered for further work.
10. User D would like a category tree to locate questions rather than the existing system that uses a single level of question categories. This feature was not included in the next version of the software, used for evaluation three. Adding subcategories would require a significant structural change to the existing architecture.
11. User D would like the questions already asked, to be recorded in the medical record as a list. This feature was not included in the next version of the software for evaluation three, but could be considered for later versions.
12. Both users found that if a text description of a hypothesis is deleted, the hypothesis record is not deleted. Changes were made so that removing the text description of a hypothesis leads to the inactivation of that hypothesis.
13. Management options such as “explanation” and “reassurance” are not available. These were considered to be very general statements, and not central to the management of this patient. Therefore, they were not included for the third evaluation cycle. Nevertheless, their inclusion would be reconsidered after

additional feedback from other users, including medical students and general practitioners.

14. Standardized information on Type 1B Glycogen Storage Disease was not available from within the application. An information sheet was prepared (see Appendix H on page 261).
15. User D did not believe the prescription of Gemfibrozil was a critical action. Given this user's extensive experience in lipid management the prescription of Gemfibrozil was re-categorized as a relevant action.
16. Review of the questionnaire by user D revealed a minor problem with question numbering. This problem was corrected.

5.5 Summary

With respect to the goal of assessing the user interface, these two evaluations cycles indicated that the software was capable of simulating a patient encounter, albeit a low fidelity simulation. As practicing clinicians, users C and D were focused on solving the medical problem rather than issues relating to the user interface itself. Nevertheless, a number of interface changes were required to improve the usability of SIMPRAC. These included more on-screen instructions on how different components should be used, as well as alterations to processes within the application. For example, suppose a user has completed all the questions they wanted to ask, when they click on the examination link for the first time, they are redirected to the hypothesis screen. Here they have the opportunity to "add" hypotheses. Having entered all their hypotheses, they then needed to click on the examination link once more, to be taken to the examination screen. Rather than having the user click on the examination link a second time, a "save" button was added to the hypothesis screen. After entering their last hypothesis,

clicking on this button adds the last hypothesis to their hypothesis list, and redirects the user to the examination screen, or other screen as appropriate.

As expected, the biggest interface problem related to the question and answer interface. The main issue was that the interface did not recognize all the questions made by each user. This arose from specific questions not being held in the database, the software implementation being unable to cater for minor misspellings, and the inability of SIMPRAC to recognise questions asked in different ways but still representing the same concept. While the question and answer interface was not optimal, it was not a major focus of this thesis. Despite its limitations, it was considered adequate to enable further exploration of support for learner reflection in a simulation environment.

With respect to the review component, based on feedback from the first four think-aloud sessions, and with developing ideas on how reflection could be encouraged, the software was modified to include three new features. Firstly, the review component was modified so that it opened automatically at the end of each consultation, thus prompting the user to review their activity. Secondly, an additional element was added that forced the learner to review and categorise their own actions and selections as being critical, relevant, or not relevant to the diagnosis and management of the patient (see Figure 25 on page 65). This classification process must be undertaken before users can compare their activity to their peers or an expert. In contrast to the graphical review component, which was designed to provide feedback and encourage reflection, this component is more directive in making the learner undertake a reflective process. Thirdly, the review charts were modified so that learners could call up specific details on which of their questions, examinations, investigations, and management options had been considered

critical, or relevant by the case author, as well as which actions they had chosen that the case authors had not considered relevant (see Figure 30 on page 70).