Chapter 7 Discussion

This chapter reviews what SIMPRAC may indicate about the support of learning through reflection, in the context of medical students and practitioners, working through a series of simulated consultations involving the diagnosis and management of chronic illness. When this project commenced in 1997, there were very few web-based simulations available (Hayes and Lehmann, 1996). Those that were available lacked sophistication, did not consider the management of chronic illness, and did not explicitly support the learning modes we were interested in. Proprietary software of varying levels of sophistication was available at the time. For example, the product known as DxR (Myers et al., 2001), available on CD-ROM, provided a rich interactive experience, and had just been made available over the World Wide Web. Nevertheless, it was proprietary software and not open to modification, so components could not be added to explore support for reflection. Other systems, such as the text-based system developed by the National Board of Medical Examiners (Clyman and Orr, 1990), did enable medical management to be implemented over multiple consultations. However, this software was used for certification purposes, was again proprietary in nature, and was not available for modification. Furthermore, with the improvements in technology, and the ability of web-based technologies to be widely available in time and place, a web-based approach was considered most desirable. In order to address these design requirements, a project was undertaken to develop the software “in house”.

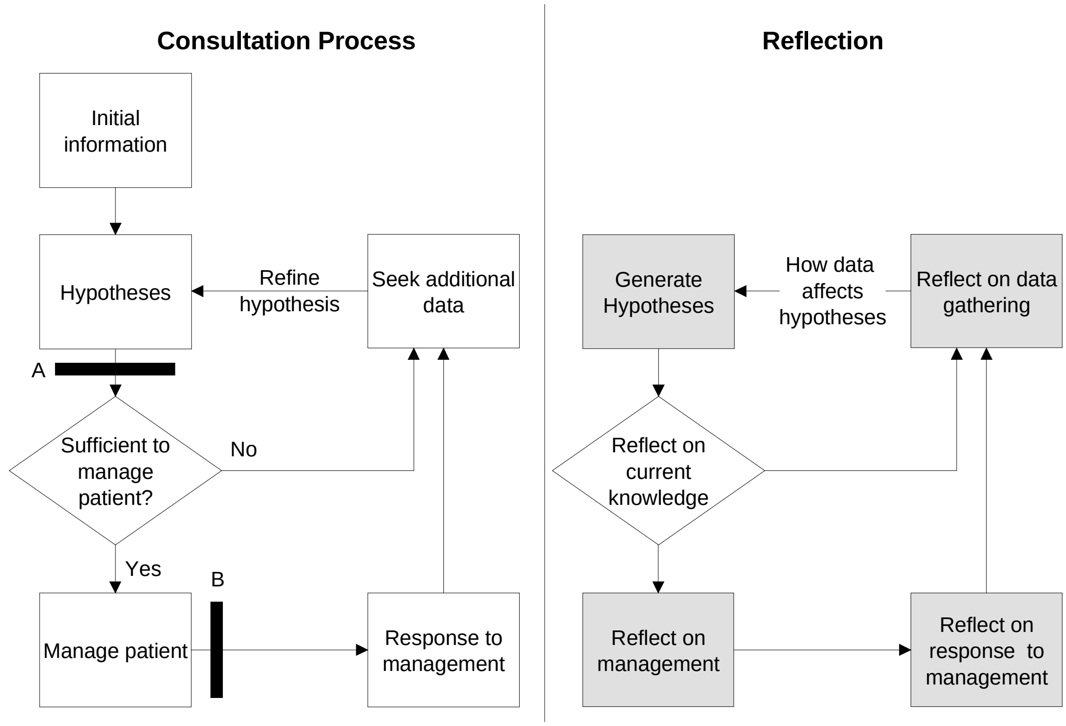

Figure 62: Model of the consultation process. Reflective processes explicitly supported in SIMPRAC are shaded grey. Steps A and B indicate where the consultation process is interrupted to promote reflection.

As discussed in Chapter 5, SIMPRAC is based upon a relatively simple model of the consultative process. This model is reproduced in Figure 62. Most of the diagnostic and management interface has been developed to support these processes, and enable users to diagnose and manage a virtual patient. In addition to those elements included strictly for interacting with the patient, a number of other elements have been incorporated, with the goal of supporting reflection. This chapter begins by discussing how well SIMPRAC has been able to provide a suitable environment for enabling users to interact with a virtual patient. It then goes on to discuss those issues relating to the support of reflection. In particular, the questions, is SIMPRAC able to support reflection, does SIMPRAC encourage reflection, and what additional factors may be necessary to improve support for reflection, are discussed.

7.1 SIMPRAC as an authentic activity

Constructivist approaches to adult learning emphasize the need for an authentic learning activity (Merrill, 1991, Spiro et al., 1991b). In the medical context, this has traditionally been undertaken though bed-side teaching. With the changing economic and social climate, it has become more difficult to rely solely on this resource, and various simulation-based approaches have been explored (Bergin and Fors, 2003). Some of the most realistic simulations used in research and teaching have involved the use of trained actors (Bergin and Fors, 2003, Elstein et al., 1990). However, this approach has been associated with difficulties in actor availability, reliability, and standardisation, as well as the high cost of implementation and maintenance (Hubal et al., 2000). While computer-based simulations have yet to reach the sophistication of trained actors, they have reached a level at which effective learning can take place (Devitt et al., 1998). Therefore, when designing and evaluating SIMPRAC, the goal was to produce a simulation that would enable effective learning to take place, rather than to produce a perfect simulation, *per se*.

With this particular software application, there are four major components corresponding to the typical activity undertaken when diagnosing and managing patients. These are history-taking, physical examination of the patient, requesting and reviewing investigations, and the development of a management plan.

7.1.1 History Taking

In an attempt to minimize prompting of the user, an attempt was made to enable users to enter their questions as free text or keywords. The entered question, phrase or keywords are compared to the list of available questions, and a list of possible matching questions displayed to the user. By clicking on the question most closely subserving the user's intended inquiry, the user can see the patient's response. As an alternative, users can also select the available questions from a series of categories. There was a wide variation in how the question interface was used. Some individuals such as ST07 and GP5 only asked their questions from the category lists. In contrast, ST04 asked all their questions using the free text method. It should be noted, however, that this user also asked the fewest questions, second only to GP4. This general practitioner asked three of their four questions in the first consultation, using the free text method. Over all, the free text method of asking question was only able to retrieve a suitable match approximately fifty percent of the time. From observations during the think-aloud sessions, most users started by trying to use the free text method of asking questions, but became frustrated when their questions were not recognised, and started using the category lists. This was exemplified by ST10 who started by trying to ask a number of open-ended questions but after a number of attempts then said, "it might be easier to see what it will allow me to ask". From Questions 3, 4 and 5 of the questionnaire (see Appendix C) it is apparent that, while most people would have preferred to ask their questions as free text, many found it difficult to use. Users were equivocal about how well it recognised their questions. Furthermore, it is unknown from the questionnaire results whether those people who preferred to use the category list mode, held this preference because the software was not able to understand their free form questions effectively.

The inability of the software to find a suitable matching question arose from two distinct causes. The first was related to the relatively unsophisticated technology used to parse the questions. The software used a free text indexing system (Apache Software

Foundation, 2002) and there was no attempt at true natural language understanding. The second cause, and probably the major cause, was the fact that many of the open-ended questions asked by GP2, GP3, and ST10 were not available in the corpus of questions held in the database. This highlights the problem of allowing users to use natural language, in that it opens up the universe of all possible questions. With a suitably sophisticated disease model, integrated with a natural language processing engine that could relate the semantics of each question to the knowledge of the disease process, it might be possible for users to get more appropriate responses to their questions. However, such technology is not readily available, and development and maintenance of the necessarily sophisticated models of disease processes, would make authoring of additional medical cases difficult. This is especially true if they were in different medical domains. One of the reasons for using a relatively simple model was to enable SIMPRAC to be used as a template for many different areas of medical practice. Rather than using true natural language processing, one could just add more questions to the database. On the other hand, this too has its problems. To cover all the possible questions that could be reasonably asked by users, the question database could potentially become very large. This approach was undertaken by the developers of one web-based interactive patient, and the number of cases available has remained small (Hayes and Lehmann, 1996). Such a large database of questions also makes authoring cases quite difficult, as each question needs to be reviewed to ensure that only coherent responses, on behalf of the patient, are made. This problem is accentuated further where multiple consultations and multiple patient outcomes are possible. Having noted the above issues, the following paragraphs explore how the questions and answers could be contextualised to the medical case under consideration, and enhance the user experience.

The use of the lists of questions from different categories was not without its problems. Firstly, some users had difficulty finding a suitable category. For example, ST10 was looking for questions regarding the patient's family history, and was frustrated when he was not able to find a category with this name. There was a single question regarding a history of family illness and this could be found in the medical history category. A second problem arose from the fact that the questions within the question database were not case-specific, and the list of questions relating to the presenting complaint was very long. EX2 stated,

> *"The lists of questions are too long. I wanted to read through the list to ensure that I haven't missed the most appropriate / relevant questions, but by the time I finished reading, I have already forgotten what questions were there at the top."*

Using non case-specific questions was an intentional design decision, made to minimise prompting of the learner, as well as to make it more readily generalizable to other cases in different domains. However, many of the questions were not even remotely relevant to the case. This arose from the fact that early iterations of the project were developed with a focus on food allergies in children. Thus there were questions relating to a child's history of allergy, illness, and immunization, even though the case, used for the later evaluations, was based around an adult patient with no children. Again contextualising the questions to the medical case would be a desirable feature. On the other hand, although constraining the available questions does have its merits, it reduces the authenticity of the simulation, and conflicts with the desire to avoid leading the learner. Minimising prompting was one of the reasons free text question entry was included in the design of SIMPRAC.

While the questionnaire results are fairly neutral with respect to how much difficulty subjects had in asking the patient questions, the design of the evaluation may have led to a more positive assessment of the user interface than would have been made under less controlled conditions. With the simplified think-aloud there is some interaction between the person making the observations, and the person interacting with the software under study. In the case of this evaluation, the person undertaking the observations was also the author of the software and case. From time to time during the think-aloud, the user would become stuck, and not know where to find a question or a feature. Having the software author available meant that he could intervene in the process. For example, when ST10 was getting frustrated at not being able to find the question regarding family history, the author was able to indicate to the user under which category that it could be found. This intervention probably led to a more positive assessment than would have been the case if the author had not intervened. In a similar fashion, there was some information, such as the patient’s occupation, which was not encoded in the database, and a question relating to this would not have been found, no matter how hard the user searched. Nevertheless, this evaluation process does enable user’s interaction with the program to continue, having identified the problem. In this way multiple issues can be identified and addressed from a limited number of sessions.

Based on the above information, it is apparent that more work is required with respect to the history-taking interface. Having two methods of use is not unreasonable, as it enables users of the software to utilize alternative methods of finding the questions that they seek to ask. Having a free text mode seems to make people believe that any question can be asked, although this could be countered by appropriate training and experience, so users are aware of the limitations. Instruction on the use of keywords would also enable more effective use of the free text entry method of choosing questions, by limiting the possible matching questions to a shorter more directed question set. This would answer the criticism of EX2 (page 173). The existing text-processing engine probably needs to be modified to allow for minor spelling variations. Entering “smoking” and “smoke” should return similar values. The question database may also need to be revised so that the questions available in any single case are more focussed on the case of interest. That is, there might be questions that are considered not relevant to the diagnosis and management of the patient but at least they would be sensible in the context of the patient. A limited number of open-ended questions may also need to be added to the corpus of questions.

7.1.2 Physical Examination

The examination component was not a new design and embodied elements from similar web-based applications (Myers et al., 2001, Hayes and Lehmann, 1996). In contrast to other systems, which use patient photographs, this software uses a human silhouette, against which various examinations can be performed. This approach was taken to reduce the amount of work in authoring cases by eliminating the need to produce an image map for each new photograph. Once familiar with the tools and the various views of the body, users appeared to have no difficulty using the interface. Nevertheless, the scores on the questionnaire were either neutral or only slightly favourable, while ST07 and GP3 gave it quite low scores in terms of usability. As a result of there not being a photograph of the patient, users were not able to easily obtain a visual overview of the patient. During the think-aloud sessions a number of users were noted to use the inspection tool extensively, thus contributing significantly to their activity scores. This was particularly true for a number of medical students, as well as EX1 (See Figure 44 on page 143). These users used the tool over a large part of the body in an attempt to obtain the overview they desired. An alternative to having a mapped image for each case might be to include a photograph of the patient as part of the introduction to the case. Such an alternative might be able to provide improved satisfaction with this component, and should be included in any further evaluations. This was not done for the current case because the case author had not seen a patient with this condition for a number of years, did not have a photograph of a suitable patient, and would have had to negotiate the considerable ethical difficulties involved in obtaining consent from such a patient. One possible solution to this impasse might be to generate morphed pictures of normal, informed, or paid, volunteers.

7.1.3 Investigations

Based on the results of the questionnaire, most people indicated that they found it easy to order investigations. The average score for this question was greater than 4 for the students, general practitioners and specialists. On the other hand, the list of blood tests, with 392 entries, was excessively long, and most users used the search facility to find the test they required. The think-aloud observations did indicate that some training was required before the user could use the system effectively but once trained they could generally find the tests they were looking for. There were instances where they could not find the test they wanted to request. This was a result of either:

1) The test was not in the database. For example, a request for a dual x-ray absorptiometry scan to assess bone mineral density.
2) Use of synonyms or abbreviations. For example, uric acid instead of urate, or LFT instead of liver function tests.

The addition of a synonym list would be a simple change that would increase the ease with which users could find the tests they desired.

Due to the repetitive nature of requesting diagnostic tests in this case, where the patient was monitored, largely by the results of their biochemistry, two students suggested that a function should be established that would enable panels of regularly used investigation to be ordered more easily. Based on their experience with existing pathology systems, a number of users also requested that abnormal values be highlighted. Another request was for a cumulative view that enables results from previous consultations to be viewed at the same time as the most recent results. At the time this project was commenced, there were very few systems available for electronic ordering of pathology and other tests. In more recent years, a number of products have been developed that are either part of laboratory information systems (PJA Computer Consultants, 2004), standalone programs (Atlas Medical Software, 2003), or part of point-of-care clinical systems (Health Communications Network, 2003). However, they are not standardized, and the software is proprietary and not open for general use. Pathology reports tend to have a standard format, and software is available for reporting results via a web browser. However, being proprietary, this software is not open for use within SIMPRAC.

7.1.4 Management

Most of the medical students indicated that they thought that it was easy to order the various management options. Similarly, most general practitioners and both experts also indicated that they thought it was easy to select management options. In contrast, GP3 and GP6 did not think it was easy to use. However, this largely stemmed from the fact that they were not able to order the management options they desired. Based on observations during the think-aloud experiments, most users had some difficulty finding the options they desired. This was largely due to the fact that the management options were limited and relatively specific to the lipid management aspects of the case (see

Appendix J on page 276). These issues are discussed in more detail below as they represent limitations in the authoring of the case rather than limitation in the design of SIMPRAC, which can include any number of management actions.

Several users wanted a finer level of interaction with respect to management. For example, ST10 would have liked to have been able to prescribe a diet high in complex carbohydrate, but low in simple carbohydrates. While the option to use uncooked cornstarch was available, this did not match the concept the user had in their mind. Furthermore, the management options were only available from a free text enquiry so it was not possible for users to scan a list of available options. SIMPRAC was purposely structured in this way, so that users would not be prompted on the available management options or options that would be most relevant to the case under study. In a similar fashion, GP6 wanted to be able to negotiate different levels of intake of cornstarch with the patient. This particular general practitioner was concerned regarding the difficulty with which this patient was able to comply with cornstarch therapy, and wanted to negotiate with the patient regarding their intake of this substance. They also wanted to explore alternative methods of preparation, with the goal of making the cornstarch more palatable.

The observations above, regarding the level of detail available to the user in their management of the patient, represents a similar problem to the question and answer interface. By having a free text entry, users expect that all possible options are available and that all natural language interaction will be understood. Within the limits of existing technology, constraints need to be placed on the level of detail that is available to the user. More detailed options are only needed for those core areas that are the major focus for learning. In the current case, it was only desired that the users learn the broad principles of management (see expectations of the user on page 46).

Most of the general practitioners, including GP2, GP3, GP4, and GP6 wanted to explore additional issues such as opportunistic screening. Given the age of this patient, most of the general practitioners wanted to discuss cervical cancer screening and the performance of a Papanicolaou smear. GP2 commented, “Doesn't recognise all the elements of a consult. Reflects hospital/specialist ‘silo’ thinking compared to GP ‘green fields’”. Similarly, GP4 thought the case was “one dimensional”, as they would have expected more problems. The unique characteristics of family practice compared to the usual medical paradigm are discussed by Shapiro and Talbot (1991). They state, “the biopsychosocial assumptions inherent in ‘care for the whole person’ stand in potential opposition to the dominant biomedical paradigm”. This difference in approach between general practitioners (breadth) and specialists (depth) was exemplified by EX2. EX2 is a specialist in lipid management but has no experience in the management of Type 1 GSD. As evidenced by Figure 46 on page 144, this particular user had a logical stepwise approach to this patient’s dyslipidaemia. Nevertheless, they appeared to fail to consider the patient’s glucose control. It should be noted that, the prevention of fasting hypoglycaemia, by the use of uncooked cornstarch, is the treatment of choice for Type 1 GSD, and leads to improvement in dyslipidaemia without the need for hypolipidaemic agents such as Gemfibrozil (Chen and Burchell, 1995). Nevertheless, hypolipidaemic agents may need to be considered where glucose control is suboptimal, or where additional factors may be influencing the concentration of cholesterol and triglycerides in the blood.

From these observations, the number of management options certainly needs to be increased. How these are rated, will depend on who are the intended users of the software. If general practitioners are expected to be using the software as part of their continuing professional development, then the development of cases needs to be done in collaboration with this group, to ensure important components, relevant to their world view, are included. This also has relevance to how the different management options are rated, and how they are expected be used in the review components. For example, inclusion of a Papanicolaou smear as a relevant investigation might be suitable if general practitioners are being targeted, but less relevant if hospital resident medical officers are the intended user group. On the other hand, one must consider carefully, what is the intended purpose of the simulation. Most of the general practitioners, based on observation and their responses to the questionnaire, thought that the purpose of the software was to simulate a general practice consultation. In fact, the case was chosen to explore the diagnosis and management of hypertriglyceridaemia. It was hoped that users would consider both primary and secondary causes of hypertriglyceridaemia, and manage the patient accordingly, when they learned that the patient had a rare secondary cause. It certainly wasn't planned to simulate all aspects of a general practice consultation. That is to say, the author and the user group had manifestly different conceptions of the purpose of the simulation. For effective teaching and learning, these must be better aligned. For this to occur, either the changes mentioned above must be painstakingly executed, or the aims of the simulation must be made explicit to the user group.

A further issue related to comprehensiveness was highlighted by EX2. At the end of the third consultation, EX2 wanted to prescribe the plant sterol containing preparation called Basikol. As use of this as a treatment modality was unknown to the author, it had not been included as a management option. This highlights the fact that medicine and medical knowledge is continually changing, and simulations and other entities used to assist learning, need to be kept up to date, and revised frequently. In this respect, using a web-based or client-server model has its advantages, as the knowledge repository is held in just one place and only has to be updated once. Nevertheless, the architecture still needs to be capable of being able to accommodate these changes and updates without having to completely re-author a case. To some extent, this suggests that there will always be a place for human interaction and review. On the other hand, deep approaches have been taken such as the GUIDON project, in which a dialogue-based learning system was based on an existing expert system, with a comprehensive disease model regarding the treatment of infectious diseases (Clancy, 1987). In this way, the learning system is kept up to date automatically. Nevertheless, this will only occur so long as the expert system is maintained and updated to accommodate new information. Furthermore, and as mentioned previously, using such a deep model, with existing technologies, tends to limit the scope of the learning environment.

7.1.5 Multiple Consultations

The multiple consultations and streams define a finite series of states (Figure 21 on page 60). Transitions between these states, is governed by the consultation and management options selected by the user. From the questionnaire, most of the users believed that having multiple consultations made the simulation more realistic but did not make it more tedious. In particular, ST05 noted during the think-aloud session that medical students seldom get to see returning patients. This is due to the fact that the time spent with any one medical team is relatively short. ST05 saw the simulation as a way of getting experience with patients that return for review.

GP6 did comment that, “the patient didn't seem to progress with the visits.” Little progression in the patient’s status was also seen for ST05, ST09, and ST10. The stream pattern for these students was [1,3,3,3] for consultations one to four. By reference to Table 8 on page 135, it can be seen that most other users had a pattern of [1,4,4,6] or [1,1,4,4]. These latter two patterns indicate that the user has taken the preferred approach of prescribing a fibrate to lower the triglyceride levels at the end of the first or second consultation, respectively. In response to this action, the user would have seen a fall in the triglyceride concentration from a level of around 12 mmol/L to 4 mmol/L. Note that the upper limit of normal for triglyceride is approximately 2.0 mmol/L. Interestingly, some students did not measure the triglyceride concentration at the first consultation, but requested it during the second consultation, after seeing that it was a critical test during the review process. Based on the fact that the level was twice the upper limit of the reference interval, they assumed that the treatment had been ineffective, as they were unaware that prior to fibrate therapy, the triglyceride concentration had been three-fold higher.

The stream patterns of [1,3], [1,3,3], and [1,3,3,3] taken by ST03, ST10, and ST09 respectively, indicate that these users prescribed an HMG CoA Reductase Inhibitor (statin) at the end of the first consultation, and that this treatment was maintained in subsequent consultations. In this case, and reflecting the real situation, prescription of a statin alone was associated with a significant but suboptimal fall in serum triglycerides, as well as a significant fall in serum cholesterol. While the fall in triglyceride was suboptimal, it was sufficient to prevent the onset of pancreatitis (stream 5 in consultations 3 and 4). Furthermore, the use of statins had been scored as a relevant option. Thus, students who took this path knew that they had instituted a relevant therapy, but were not aware that it was suboptimal. As the patient was no longer at risk of pancreatitis, there was no feedback to the user (other than comparison to their peers or an expert) that an alternative approach should be tried, and the users just maintained the same treatment regimen. This illustrates both a limitation of using natural feedback, and the potential usefulness of being able to compare activity against a peer group or an expert. Comparison with others, associated with reflection on the differences between the activity of the user and the activity of others, may lead the learner to consider alternative management strategies, leading to better patient outcomes. Nevertheless, there are caveats to this latter approach that are discussed later.

In both of situations outlined above, the attempt to make the simulation more realistic may have overshadowed the educational objectives. An alternative approach, as suggested by User C, would be to include a case summary that includes all information that should have been elucidated. An alternative or complementary approach to having a case summary might be to include a scoring system with more than three levels. In this way, those items considered more optimal or appropriate (in this case, the use of a fibrate), would have a higher score than the less desirable items (in this case, the use of a statin).

This same problem, where the system provides feedback to the user on what has been missed, but does not provide any explicit information on why it was important that something should not be missed, or what they would have learned about the patient if that particular item had not been missed, is further exemplified by GP4. This practitioner hypothesized that the patient had hyperlipidaemia as part of the metabolic syndrome characterised by hypertension, obesity, insulin resistance, and hyperlipidaemia. This practitioner completed all four consultations without ever knowing about the history of GSD, as she did not ask the patient about their past medical problems. Again, it may have been possible to avoid this situation by providing a summary of the information the learner was expected to have elicited, at the end of each consultation. Such an approach has been used extensively in a class of intelligent tutoring systems, known as cognitive tutors. These systems are based around a cognitive model of the task, or procedural knowledge to be learned (Anderson et al., 1995a). Anderson (1995a) has reported that:

1) These tutoring systems improve the rate of learning.
2) The best interaction is provided by immediate feedback consisting of short and direct error messages.
3) The tutors work best if they are presented as non-human tools rather than an emulation of a human tutor.

While the educational benefit of this approach has been demonstrated, they have primarily been used in the, arguably, more restricted domain of mathematics and computer education (Anderson et al., 1995a, Koedinger, 1998). Even so, the development of the cognitive models have been difficult and costly (Koedinger et al., 2004). In the less well defined domain of medicine, these difficulties are likely to be amplified many times. Furthermore, these systems have been focussed on mastery of procedural knowledge, rather than the ability to develop deep knowledge and transfer, which are the goals of this project, Nevertheless, these systems do indicate that provision of rapid feedback on what the student knows, and what the student doesn’t know, does improve learning.

The other major issue with the current method of encoding multiple consultations is the combinatorial effect. With the addition of each factor that may affect the outcome of the patient, the number of possible outcomes increases as a product of the number of factors, for each additional consultation. With each additional state, the history, examination and investigation results need to be reviewed to ensure that they are consistent with the current state of the patient. This makes authoring a case of more than two or three consultations difficult, time consuming, and prone to error. Furthermore, this state-based approach, also means that there are a number of patient states that need to be coded that few, if any, users will use. This problem is not easily solved and reflects the true nature of medicine and other complex systems, where multiple interacting factors can lead to a large number of possible outcomes.

An alternative to using a deterministic state-based model is to use a probabilistic model where the outcome has a realistic chance of occurring (Melnick, 1990). However, this approach would need a more sophisticated disease model, which, as mentioned above, would probably limit the portability of the application to different disease states. Moreover, while it may add to the realism of the case, and may reduce the cognitive load required to map the outcomes compared to using a deterministic approach, it doesn’t solve the combinatorial problem. Furthermore, if users were to be allowed to repeat a consultation, use of a deterministic model is more likely to lead to consistent patient outcomes.

Another major problem resulting from the combinatorial effects was the large opportunity for unforeseen interactions and behaviours. For example, it was not anticipated that having been told that high triglyceride concentration was the reason for the patient being seen, users would not measure the serum triglyceride concentration. The problem of people using safe but suboptimal therapy was also not anticipated. Another problem that wasn't anticipated was the use of incompatible treatment. Both statins and fibrates increase the risk of myositis, although statins are more likely to be associated with this side-effect (Badewitz-Dodd, 1996). Use of both medications, however, increases the risk synergistically. When encoding the various states, provision was not made for an adverse reaction if these two drugs were prescribed together. This problem highlights the cognitive load that is placed on the author when building a case, and indicates that cases need to undergo iterative improvement. These improvements include such things as unpredicted or forgotten elements, new investigations, or new treatment modalities.

While the combined use of statins and fibrates by medical students and general practitioners should be discouraged, judicious use, depending on the context of the individual patient, may be a reasonable approach for a specialist. Indeed, EX2 did use this as a combination. This has even more implications in the review section of the application. One of the goals of the review section was to enable users to compare their activity and actions to a peer group and to an expert or specialist. This was premised on the idea that the specialist would be using the most appropriate treatment and would be a "model" for the other users. However, as demonstrated in this case, one specialist instituted management that should only be initiated by a specialist, which is not something that would be desirable for other non-expert users to emulate.

As previously mentioned in Section 4.2.4, the current algorithm for managing the patient state transitions does not allow for direct interactions between management actions. The increased risk of myositis with the combined use of fibrates and HMG CoA Reductase inhibitors is an example of such an interaction. Use of a rule-based system, in addition to the existing mechanism, could be used to capture this exception and place the patient in a state where myositis or rhabdomyolysis (break down of muscle) is a consequence of using this combination of therapeutic agents. Moreover, it is conceivable that an expert user could, at the end of a consultation, reflect on those treatments or actions that they would not recommend for non-expert users. These situations could then be registered, so they could then be highlighted to subsequent users.

A further issue, concerning authoring multiple consultations, relates to how to get the expert activity into the database. At the present time, this is captured by having an expert use the simulation and simply recording their activity. However, the expert will only traverse a single path through the case. This is also true of any other user of the application. Presuming the specialist traverses the consultations using a near optimal approach, there will be some streams where there is no expert comparison group with which the user can compare. For example, ST08 had a stream sequence of [1,1,5,4] which means that she had not prescribed any effective lipid-lowering therapy for the first two consultations, and so the patient entered stream 5 and was reported to have suffered an attack of acute pancreatitis secondary to their elevated triglycerides. As no other user, and no expert, had entered this stream, there was no comparison group available. With the current architecture, the only way to ensure that data from an expert is available for every state within a given case, would be for the expert to be presented with the patient in each of the possible states. For example, as illustrated by the state transition diagram for the biochemistry case (Figure 21 on page 60), for all states to be covered, an expert would have to undertake 17 consultations. While it may be possible to do this, it is not really practical.

One approach to resolving this problem would be to make the cases, in some respects, discontinuous. That is, at the end of each consultation, the outcome for the patient could be stated. For example, depending on the management chosen by the learner, the patient might get better, remain the same, or get worse in some way. However, the beginning of the next consultation would reintroduce the patient from a standard state, which would be the same for all users. This might have even more merit if the user could also back-track, and try alternative approaches. Such an approach would simplify the authoring process enormously. Nevertheless, it may detract from the "realism" of the case. Having noted the problems above, it is nonetheless reassuring that all users followed a "reasonable" path through the case. No users did anything that could be construed as negligent, or that would have lead to an unacceptable outcome for the patient. This is not to say that their activities did not lead to adverse outcomes. For example, not prescribing a lipid lowering agent for two consultations, led to an episode of acute pancreatitis. However, this was included as a means for providing feedback to the student, and encouraging them to reflect on the potential negative outcomes of inaction, therefore encouraging them to use an effective agent.

7.2 SIMPRAC: Feedback and Reflection

From the literature, medical problem-solving seems to take place in at least two different modes. Where the clinician has past experience with a similar case, they use illness scripts that contain information about the clinical context of the disease and its consequences (Schmidt et al., 1990). Where the clinician encounters an unfamiliar problem, then an alternative, hypothetico-deductive approach is taken (Elstein et al., 1990). This latter approach involves the user generating a short series of hypotheses regarding the patient then, exploring these through the process of history-taking, physical examination, and the use of various investigations. Reflection has been promoted as an effective means for people to learn about a professional knowledge domain such as medicine (Schön, 1987). However, despite the promise for effective learning, teaching methods involving feedback and reflection have generally been underutilized in primary and continuing medical education (Branch and Paranjape, 2002). Issenberg (Issenberg et al., 1999) noted that,

> *"The most important identifiable factor separating the elite performer from others is the amount of deliberate practice. This includes the duration of practice and the amount of effort. Deliberate practice needs to be a well-defined task with an appropriate level of difficulty for the individual. There should also be informative feedback and opportunities for repetition and correction of errors."*

Consistent with the reflective practice processes listed by Imel (1992) (see page 10), SIMPRAC was specifically designed to incorporate elements that encourage and support reflection. A discussion of where reflection can be supported, and how each of the reflective elements in SIMPRAC relates to the consultation model, can be found in Section 1.4 of Chapter 1 (page 15). The reflective elements relating to the consultation model are shown in Figure 3 on page 39, and are listed below:

1) Asking the user to review their hypotheses at the end of each diagnostic stage (Hypothesizing).
2) Asking the user to review the relevance of each of their actions at the end of each consultation (Question what, why and how one does things).

3) Enabling the user to review their actions compared to an expert, and their peer group (Ask what, why and how others do things, comparing and contrasting, asking for others' ideas and viewpoints).
4) Providing natural feedback on the patient's response to the user's management orders (considering consequences).

As illustrated in Figure 62 on page 169, these reflective elements were used to interrupt the consultation at two points. Users were prompted to review their hypotheses before they performed a physical examination, requested investigation, or selected their management options for the first time in each consultation. The other elements that were included to support refection were displayed at the end of each consultation.

7.2.1 Hypothesis Generation

The hypothesis component was used by all users except ST04. Unfortunately, it is not known why this student did not want to enter their hypotheses. Users generally entered their hypotheses during the first consultation, after taking a history from the patient. Some students made some changes at the end the examination and investigation stages. However, few hypotheses were added or updated in the second or subsequent consultations. While these observations suggest that users did not see much value in this component after some initial use, there was some evidence that users were reflecting or consolidating their thoughts having completed taking a history from the patient. For example, when this screen was first displayed, GP6 sat back and said, "oh, what have I got as a hypothesis". On the other hand, since the diagnostic process is where most hypotheses are generated, and as for this case, this usually occurs in the first consultation, the fact that the hypothesis component was used less after the first consultation, is not that surprising.

EX1 indicated that he found the screen intrusive, although he understood why it was there. He suggested that it should be displayed less frequently to the user. One possibility would be to only display the hypothesis screen at times where the primary focus is on diagnosis. In the case used during the evaluations, this would mean only interrupting the user with this screen in the first consultation.

While all the students (except ST04), general practitioners, and specialist medical practitioners entered from one to four hypotheses, and users may be reflecting to generate these hypotheses, there is no mechanism currently within SIMPRAC to encourage or assess whether users are reflecting on their hypotheses. In particular, there is no structure in place to encourage users to tie their questions, investigations and so on, to their hypotheses. Should there be some process in place that encourages users to explicitly consider how a particular question, or other action, supports or refutes a particular hypothesis? While such a process may have the potential to lead to greater understanding, it was considered that making this process explicit, would be difficult to implement without it being intrusive and inhibiting use of the software. Personal experience with software such as DxR (Myers et al., 2001), where this has been attempted, has confirmed this to be an intrusive and frustrating exercise. The fact that many users did not update their hypotheses, and that EX1 specifically indicated that the hypothesis screen should be displayed less frequently, also supports the notion that making this process explicit would have reduced interest and engagement with the software. On the other hand, the potential benefit for improved reflection and learning still makes this process worthy of exploration in the future. One possibility to add this tie in to the user's actions might be to include the hypotheses with the action relevance review screen (see next section and Figure 25 on page 65). Users could then indicate which hypotheses they were exploring when asking a question, performing an examination, or requesting an investigation. While this might add some time to the reflective process, the increment may be sufficiently small to be acceptable to the user.

7.2.2 End of consultation reflection

The review components of SIMPRAC come in two parts. The first review stage is a forced review that must take place before the user can proceed to the next consultation. The user is compelled to rate the relevance of their actions, based on their current knowledge, as critical, relevant, or not relevant to the diagnosis and management of the patient. The second is a voluntary review, where the users can compare their activity to an expert or their peer group, as well as get information on what the case author considered were the critical, relevant, and non-relevant items.

Users spent much more time in the first review component, at the end of the first consultation, than in subsequent consultations. This was an expected finding, as the first consultation had a greater emphasis on diagnosis and required many more questions and actions to be performed to enable the user to gain an understanding of the issues affecting the patient. Subsequent consultations were aimed at modifying management, based on knowledge of the patient’s response to treatment and the development of additional problems, and did not require the same degree of inquiry. For example, in the case under study, the patient’s primary problem was hypertriglyceridaemia secondary to an inborn error of metabolism. Once the diagnoses of Glycogen Storage Disease and Secondary Hypertriglyceridaemia had been established in the first consultation, subsequent consultations were primarily targeted at managing the disease process.

Despite the frustration experienced by some users (Figure 53 on page 154), this exercise did seem to help users reflect on their activity (Figure 52 on page 154). Surprisingly, there was no correlation between the number of actions performed, and hence reviewed, and the degree to which people felt frustrated. On the other hand, as illustrated by ST08, lack of time spent using the various review components does not mean that the subject was not using reflection. This medical student spent less time than most users in both the stage one and stage two reviews, for all consultations (Figure 40 on page 140 and Figure 41 on page 140). Yet this same student, at almost 107 minutes, used the application longer than any other user. This student took almost 70 minutes to complete the first consultation (Table 8 on page 135, and Figure 39 on page 139), again longer than any other user. During the think-aloud session, ST08 was constantly rephrasing the patient's issues based on the information that was available. For example, after asking the very first question, "why have you come today?" and learning that the patient had come for management of their hypertriglyceridaemia, ST08 stated out loud that she needed to:

1) Explore what other factors had lead to the problem.
2) Explore additional cardiovascular risk factors.
3) Explore any potential consequences of the lipid abnormalities.

Further activity was then directed to addressing these issues. In the schema of Schön (1987), this student appeared to be very adept at *reflection-in-action*. Interestingly, the stream pattern for this student [1,1,5,4] was quite different from most other users. Based on the information sheet on Type 1 Glycogen Storage Disease (Appendix H), this student tried to improve the patient's metabolic control by prescribing uncooked cornstarch, and a diet low in saturated fat. Only after this was unsuccessful, and the patient experienced an episode of pancreatitis, did she prescribe a specific lipid lowering agent. While this might have been a suboptimal path, the reasoning behind it was sound. In summary, the think-aloud observations suggest this student is very good at *reflection-in-action*. It remains unclear how much this student was employing *reflection-on-action* or *reflection-for-action.* ST08 did not use the tools intended for this purpose to the same extent as their colleagues. This raises the question as to whether these two modes, *reflection-in-action* and *reflection-on-action*, are inversely related. On the other hand, while actual performance may be improved through active *reflection-in-action*, learning is greater if there is *reflection-on-action* and reflection on reflection (Schön, 1987, Ruth-Sahd, 2003).

These observations demonstrate the value of using a think-aloud experiment to evaluate SIMPRAC. Only by having one-on-one observation was it possible for ST08's *reflection-in-action* to be observed. Without direct observation, it would not have been possible to account for the time this user spent on the case. These observations also suggest that *reflection-on-action* may be better supported, at least for ST08, by having interface elements that more actively focuses the learner on the critical concepts to be learned. Furthermore, and as discussed later, reflection may also be encouraged, by ensuring learners are familiar with reflection as a powerful way of improving learning, in which case they may be more willing to use existing tools designed to help them in this activity.

In contrast to ST08, ST10 did not appear to spend a lot of time during the case reflecting on the presenting issues, but nor did he spend a great deal of time using the review components to reflect on past actions. Unlike all other students and general practitioners, this student had substantial previous exposure to patients with Type 1B

GSD, and was fully aware of the dire consequences of the infective complications suffered by these patients. However, despite being informed that the infective aspects were being managed by an Immunologist, this student continued to be primarily concerned with the possibility of infection. On the other hand, the case was authored by a medical specialist, and as observed by the general practitioners, was focused on a single major pathology. Given the possibility of life-threatening infection in these patients, it is not unreasonable to make it a higher priority for attention than envisaged by the author of the case. However, this appears to have been done at the expense of considering the metabolic derangements, thereby substituting one ‘silo’ for another. These observations indicate the user was not aware of their omissions, and while these could have been identified by more thorough use of the review tools, this did not take place. One solution to this problem, based on the cognitive tutor approach, would be to provide more explicit feedback on the omissions. From a learner reflection point of view, one could also ask the user to reflect on why these omissions were important, and how they relate to the patient as a whole.

Clearly, the issues to be focussed upon need to be made clear to the learner. As discussed in Section 7.1.4, when authoring the case, the learning objectives need to be identified, and the objectives of the learner must be aligned with those of the case author. This is not to say that the case author dictates to the learner what they will learn. Rather, knowing what the user wants to learn, the case author can develop a case that highlights the important issues. Another implication of the above observations relates to the power of using multiple consultations. That is, the natural feedback provided by patient outcomes can be used to highlight the factors important to the correct diagnosis and management of the patient. For example, in the case under study, failure to use an effective lipid lowering agent led to an adverse patient outcome, which was relayed to the user at the following consultation. All users, when faced with an adverse event, took appropriate action to ensure a better outcome.

Schmidt et.al. (1990) in their theory of medical expertise, promulgate the idea that practitioners develop “illness scripts” by first being exposed to a sentinel case. These scripts are further developed and refined through experience with similar cases. These scripts, “contain relatively little knowledge about pathophysiological causes of symptoms and complaints but a wealth of clinically relevant information about disease, its consequences, and the context in which it develops” (Schmidt et al., 1990). In this theory, memories of past patients influence the management of future patients. This seems to have been the case with ST10, where past experience with patients with Type 1B Glycogen Storage Disease has influenced the student’s approach to the case under study. Interestingly, this case was chosen because the condition was uncommon, and unlikely to have been seen by any of the students or medical practitioners. It also indicates that the sentinel case can have a profound, and perhaps overly great, influence over future practice. It would be interesting to explore how future real or simulated cases are able to modulate a practitioner’s illness scripts, especially those that may be suboptimal following exposure to a case that is highly memorable, as a result of an individual’s idiosyncratic experience.

The experience of ST08 and ST10 highlight the fact that, while this application provides tools to support reflection, it does not attempt to convey the importance of, or teach reflection. It is known that people do vary in their natural degree of reflectivity (Wong et al., 1995) and that some may not be aware that they reflect even when they do (Boyd and Fales, 1983). However, this is not necessarily an either-or-situation, and individuals may engage in different levels of learning, depending on the task at hand, and the goals of the individual (Moon, 2001). Wong et.al. (1995), in their review of reflective journals, describe three types of individuals. There were the non-reflectors, who reported only what happened. There were the reflectors who considered contextual aspects of their experience. Finally, there were the critical reflectors, who examined themselves and the experience in a critical manner.

When the users of SIMPRAC were first introduced to the software, they were not explicitly informed that the software was designed to support reflection. Furthermore, they were not explicitly asked to use the various tools to help them reflect on their actions. Boyd and Fales (1983) observed that, “the mere naming of the process - the bringing to the consciousness of what is done naturally - is a significant aid to the use of reflective learning”. Therefore, merely explicitly informing the users that they were expected to have used the tools to reflect on their actions, may have helped some users use the review tools more systematically and effectively.

The second stage of the review facility was more flexible than the first stage, and enables users to only view information that they think is of greatest interest. All users viewed information on the questions they had asked but most did not systematically review their examinations, investigations and management options. This may indicate a lack of desire to use the tool for reflection, or it might just indicate poor interface design. With this component, each user had to view the information on questions, examinations, investigations, and management options separately. Evidence to support the assertion that, “users did not use the tool to reflect on their actions” includes the fact that, four users (two students, one general practitioner, and one expert) did not ask about alcohol as a secondary cause of hypertriglyceridaemia. The fact that this question had been missed, was displayed to all users who failed to ask about past alcohol consumption. Yet three of the four users who did not ask this question in the first consultation also did not ask about it in the second consultation. At no stage during the think-aloud sessions did these three users mention alcohol as an important secondary cause of hypertriglyceridaemia. To some extent this may reflect a lack of awareness by the individuals that, although Type 1 GSD can be complicated by hypertriglyceridaemia, other secondary causes need to be excluded. Alternatively, it might just reflect a lack of knowledge regarding the effects of alcohol intake. Another possibility is that the case was so unfamiliar, and the cognitive load so high, that these learners become wholly focussed on the primary condition.

However, there was also contrasting evidence indicating that users were using the graphical feedback to modify their actions. For example, having seen that the author had indicated that measuring the serum creatine kinase activity was relevant for a patient on Gemfibrozil, ST06 indicated that she had not realized that myositis (inflammation of muscle) and myalgia (muscle pain) were side-effects of this drug. In this instance, the user asked the author about the side-effects, based on the feedback received. Having been informed that Gemfibrozil could cause this side-effect, ST06 went on to ask about this in the next consultation (see page 116).

During the second review process, where the user is able to compare their activity to their peers, the user can also review the details of their actions during the consultation. In its current form, the user currently sees a table of:

- Those items selected by the user that the author considered were critical.
- Those items selected by the user that the author considered were relevant
- Those items selected by the user that the author considered were not relevant.
- Those items not selected by the user, yet the author considered they were critical.
- Those items not selected by the user, yet the author considered they were relevant.

The users' willingness to reflect on their omissions could potentially be improved by modifying the user interface so that users are more actively alerted to important omissions, rather than the current passive indication that a critical or relevant, question or action had been omitted. For example, in future iterations of SIMPRAC, focus could be given to those actions that the case author had considered critical, but had been omitted by the user. Interface elements could then be added that asked the user to say whether they agreed or did not agree that these items were critical, as well as asking the user to explain why they held this belief. Learning might also be enhanced by providing suggestions from the case author on why particular questions or actions are important in the context of the current case. However, the addition of such a feature would make authoring a case more difficult, as it essentially adds an additional dimension to the disease model, which considers the pathophysiology of the disorder under study. On the other hand, this might only be done for the critical elements, limiting the number of items that needed to be annotated for each patient state (see Figure 21 on 60).

Even though some users did not appear to use the tools, other users stated that they found them helpful. For example, EX1 who hadn't been actively managing patients for more than five years believed the immediate feedback was informing his questions in subsequent consultations, as some items had been forgotten. Thus the degree to which, the users found the review tools useful, may be indicative of their tendency to use reflection as a learning mode.

As mentioned in Chapter 6, a number of users disagreed with the relevance assignments given by the case author, and most if not all the users, appeared to consider that concordance of their relevance assignments with the case author's assignments indicated that they were "right", while discordance was interpreted as a statement of being "wrong". None appeared to perceive this as just a tool for reflecting on the usefulness of their questions and activities. This observed behaviour of classifying actions as being right or wrong may have been engendered by having just three relevance categories, and might be ameliorated by using a continuous scale that could be represented in the user interface as a "slider". This observation can also be interpreted as reflecting the competitive nature of medical students, and medical practitioners. Other evidence suggestive of this competitiveness and fear of failure, include statements of lack of knowledge before beginning the case, and not entering or updating hypotheses. For example, ST01, as did many other students, expressed their doubt about their ability to undertake the case, even though they were given verbal and written information that the purpose of the session was to assess the software. The fact that ST04 never entered a hypothesis, may also have been due to a fear of being wrong and appearing "foolish". Despite this, ST04 also indicated that she thought the simulation system was a good way of getting practice with unusual conditions, when there are a large number of medical students competing for access to a limited number of patients. Competitive behaviour, as suggested by the observations above, is not really surprising given the competitive process required to get into medicine, and to undertake advanced training. Yet this very competitiveness may also interfere with reflection and learning. These observations, regarding competitiveness, fear of failure, and fear of judgment, are consistent with the findings of Anderson (1995a), which demonstrated that cognitive tutors work best if they are presented as non-human tools rather than an emulation of a human tutor. Assuming the above interpretation is correct, there is clearly a need for a shift from fear of failure, to embracing recognition of errors as a foundation for learning.

It should be noted that the second stage review, that enables users to graphically compare their actions to others, is only providing feedback, and may even reinforce the competitive model. It does not necessarily ensure that the users are actually reflecting on this feedback. This distinction is very important, as Branch and Paranjape (2002) have observed:

> *"some have said that the difference between a professional and a technician is that the professional knows the larger context of his or her work and uses this knowledge for lifelong learning, as opposed to the technician, whose knowledge is limited to performing a specific task."*

They go on to say:

> *"… psychological growth occurs only when reflection is a component of an educational program. Thus reflection leads to growth of the individual – morally, personally, psychologically, and emotionally, as well as cognitively – whereas feedback tends to promote technical proficiency."*

In a similar fashion to the second stage review of this simulation, other medical, web-based, computer simulations have taken the approach of providing feedback (Hubal et al., 2000, Bergin and Fors, 2003). As such, they promote technical proficiency. However, as the literature on medical problem-solving and simulation technology suggest, this knowledge tends to be case-specific (Schmidt et al., 1990, Fitzgerald et al., 1994). While providing feedback can support a reflective process, on its own it does not ensure that reflection is taking place. Rather than just providing feedback, the stage one review component forces the user to consider the importance of their actions and may provide a greater incentive for reflection to take place, and therefore a greater chance for more generalized learning. As a consequence, one can hypothesize that placing the review screen one ahead of the chart review (as is currently the case) might stimulate more chart review than would be the case if these screens were viewed in the reverse sequence. Interestingly, as shown in Figure 61 on page 160, compared to the general practitioners, the students and experts spent relatively more time reviewing the relevance of their actions than they did reviewing the charts. This was an unexpected finding that, when considered with the fact that the total time the students spent reviewing the relevance of their activity was greater than the general practitioners (see Figure 52 on page 154), might suggest that the students were engaging in more reflection than the general practitioners. Despite the evidence that some users were using the review screens effectively, the fact that some users spent very little time in this activity, suggests that they were not reflecting deeply, if at all. The think-aloud observations supported this notion, with many individuals taking a very cursory look at the charts, without any discussion or comment. This raises the issues of the role and place of the technology.

Unlike the first stage review process, the second stage graphical review does suffer from the first start problem. That is, the first five users don't have much meaningful peer data with which to compare. Especially in the second and subsequent consultations where peers may have entered alternative streams. To ensure that only more robust data are available to users on average peer activity, it may be better only to report the peer data when more than five peers have undertaken the case with the patient in that particular state (consultation and stream). An extension to this problem also arises when the case needs to be updated in response to new knowledge, or new developments in medical practice. In this situation, previous comparative data for the peer group and the expert, may become invalid. For example, in recent years, measurement of serum Cardiac Troponin T or Cardiac Troponin I has replaced the measurement of the MB fraction of creatine kinase (CKMB) as a marker of myocardial infarction (heart attack). If a case concerning the diagnosis and management of myocardial infarction had been used before the change, users would have requested a CKMB, and this would be recorded in the comparative data. Once the change in practice has taken place, this part of the data becomes invalid. In which case, the collection of comparative data will need to recommence. If, as in this example, a new test replaces an existing test, it might be reasonable to simply substitute the new test in the existing cumulative record. However, if a test or treatment simply becomes redundant, the problem remains.

The group data comparing the time spent reviewing actions at the end of each consultations and the time spent reviewing the charts by students, general practitioners, and experts was interesting, and unexpected. These data are described in section 6.3.2.2 and illustrated in Figure 61 (page 160). There was no difference between students and the general practitioners in the time spent reviewing the charts. In contrast, the students spent almost twice as long as the general practitioners reviewing their actions. Although there were only two experts, the pattern of use by the students was more similar to the pattern of use by the experts. There was no clear reason for this observation. However, the author had the subjective feeling, in the think-aloud sessions, that the general practitioners wanted to complete the exercise as soon as possible. This reflected their busy schedules and limited time. This has great significance for continuing education, and reflection in particular, as both require time on the part of the learner for maximal effectiveness. It is interesting to speculate whether more time would have been devoted to reflection if the learning exercise had been undertaken in the evening, and away from the place of practice.

7.3 The role of SIMPRAC

As discussed above, there is evidence to suggest that some users were more reflective than others, and people do vary in their mode and depth of learning (Wong et al., 1995, Moon, 2001). Johns (1995), in his discussion on the conditions under which reflection can, “flourish or flounder” suggested that, “reflective practice always needs to be guided” (Johns, 1995). Provided learners are aware of the process (Boyd and Fales, 1983) they may choose whether or not to enter into the process, depending on what they perceive as the desired outcome. Thus SIMPRAC and, one could argue the same of any learning aid, may provide support for, and encourage reflection, but it certainly cannot ensure that it takes place. However, this is true of human-based teaching and learning. Reflection still depends on the leaner to execute it.

Successful adult education is learner-centred and engages the learner. It is active rather than passive, and relevant to the learners’ needs (Knowles, 1980). Consistent with these principles, it has been found that interactive continuing medical education most likely to lead to behavioural change requires (Davies, 1995, Eraut, 2001):

- problem-based approaches relevant to the clinical issues and involving authentic material,
- strategies that provide opportunities for participants to practice their new skills,
- methods for reminding and reinforcing clinician behaviour, as well as
- feedback on practice performance.

SIMPRAC may have a role as a tool for practicing and applying the desired diagnostic and management skills. The medical students in particular thought it had potential as a tool for practice, especially in these times of limited resources, where students may not have the opportunity to experience a large number of cases. This is especially true of conditions less commonly found in the community. In contrast to primary medical training, continuing medical education is frequently undertaken in response to on-the-job problems. A knowledge deficiency is identified and action is undertaken to obtain the requisite knowledge and solve the problem. In this scenario, a software simulation is unlikely to be the primary method for learning. However, having searched through a variety of information sources, a simulation may provide an opportunity to practice, or test one's knowledge and understanding, eventually leading to consolidation of understanding and thus to deeper learning. This might be done in relation to a new area of practice, or as revision when returning to a previous area of clinical practice. An example of a new area practice might be a General Practitioner moving from a urban practice to a rural practice where zoonoses (disease transmitted from animals) such as leptospirosis and brucellosis, are more common. User EX1 provided an excellent example of revision. As previously mentioned, having not managed patients with lipid disorders for a number of years, EX1 found that the feedback provided by SIIMPRAC helped them recall the important issues regarding the diagnosis and management of lipid disorders, especially in the context of Glycogen Storage Disease. In addition to these aspects of learning, GP2 also thought SIMPRAC might have a role in self-audit. That is, a self-evaluation tool that could be used to identify misconceptions in understanding, or deficiencies in one's knowledge.

An unexpected finding from the questionnaire was that the medical specialists indicated that SIMPRAC improved their understanding of the causes and management of hyperlipidaemia. One can postulate that this may indicate that using the application helped the experts to consolidate their knowledge, even though they had a higher baseline-knowledge than the medical students and general practitioners. An observation such as this is very encouraging with respect to continuing professional development. The case used in the evaluation was rare, even for specialists in the field. This finding suggests that while some benefit to learning is obtained by those with less knowledge, considerable benefit, through consolidation of knowledge, may be possible for those with more knowledge.

Because this software is able to record all the users' actions as they interact with the virtual patient, this technology can also be used as a tool for evaluating individuals. While most simulations are ostensibly designed for educational purposes (Bergin and Fors, 2003, Hubal et al., 2000), they have also been used to evaluate performance (Clyman and Orr, 1990, Costello et al., 1997). As reported in Section 6.3.2.4 on page 161, there were a number of questions or actions omitted by all users, when compared to the expectations of the author of the case. Interestingly, the medical specialists appeared to be less thorough in the diagnostic parts of the case. On the other hand, their management of the patient was more methodical, suggesting a good understanding of the relevant issues despite the brevity of their evaluation. This ability to quickly identify the important issues on limited information is characteristic of domain experts (Schmidt et al., 1990). Having said that, the failure of the specialists to monitor for some medication side-effects is surprising! Why this should have occurred is uncertain. It is possible to speculate, that in the absence of any history from the patient of side-effects, they believed there was little value in conducting these investigations. Alternatively, it may mean that behaviour when using the simulation may not reflect actual clinical decision making. Another possibility is that the artificiality of the think-aloud evaluation environment interfered with the fidelity of the simulation, and altered the user’s normal behaviour. Further work will need to be done to explore this observation, and understand what impact this has for using SIMPRAC as an educational tool.

With respect to patient management, the results of the evaluations did not demonstrate a large variation in the practice of the individual users. There were only a limited number of routes taken though the case and no user took any action that would have been considered negligent. By negligent it is meant, that nobody continued a course of action once an adverse effect or outcome had been identified. There was some variation in the thoroughness with which subjects examined the patients. However, this probably reflected a short-coming in the software rather than major differences in approaches to the patient. Based on these observations, it would be very difficult to develop criteria for discriminating the performance of individuals if SIMPRAC were to be used as an evaluative tool. On the other hand, one could identify some common mistakes, omissions, and misconceptions that could form part of a summative evaluation. As previously mentioned, three users failed to consider alcohol intake as a secondary cause of high serum triglyceride concentration once they learnt that Type 1B Glycogen Storage Disease was associated with this abnormality. Therefore, criteria in this case, could be whether or not a user asked about all secondary causes of hypertriglyceridaemia, or whether they asked about all the elements considered critical by the case author. With respect to the limited variation seen in the activity of the users, it should be noted, at least for the medical student group, this was a self-selected population of users. The fact that the medical students generally performed satisfactorily may just reflect the fact that, only those students who were generally competent would volunteer to participate.

Another issue, when considering using simulation software as an evaluation tool, relates to computer literacy and the software interface itself. It was also observed there was a degree of variation in computer literacy and confidence among users of the software. It is likely that these differences would impact an individual’s ability to perform well using a computer simulation. Other users, such as ST07 who stated, “I get frustrated with the technology and lose patience” found the limitations of the current version of the software interfered with the “natural” consultative process. Therefore, care must be taken to ensure user performance is a reflection of user knowledge and skill, which is not degraded or enhanced by how well someone can use a computer or a particular software package. This alludes to the more general concept of validity. That is, is the tool measuring what it is purporting to measure? To date, the qualitative evaluations have indicated a basic level of competence in terms of the consultative process itself. The students were observed to be more thorough in their approach than either the specialists or the general practitioners. With respect to management, one user did not initiate any therapy. However, this user did not complete all the consultations and therefore, did not have an opportunity to respond to the feedback. One student changed their management from a suboptimal regimen to the preferred regimen, while five students persisted with a suboptimal approach. Persistence with the suboptimal approach was probably secondary to a flaw in the feedback received, as the students were advised that the patient had improved, but not informed that a greater improvement would have been possible if an alternative management strategy had been taken. How well the users were able to integrate the information they learned about GSD and hypertriglyceridaemia is uncertain, and validation studies are necessary to see if users can abstract to new situations, having used SIMPRAC. Ideally, but with even more difficulty, studies would also need to be carried out to evaluate how well performance on SIMPRAC correlates with clinical performance and behaviour. These studies would also need to cover a number of different clinical scenarios.

While there are other computer-based patient simulations available, all of which provide some sort of feedback (Bergin and Fors, 2003, Myers et al., 2001, Bloemendaal et al., 2002), they have not all specifically attempted to provide support for user reflection. Nevertheless, many of these products are relatively mature, and have well-developed interfaces for patient interaction. Further, the approach of having multiple points of reflection is independent of the underlying implementation of the patient simulation. Therefore the SIMPRAC approach is potentially portable to existing systems or new systems, as technology improves, and more engaging and realistic environments are developed. With ongoing developments in virtual reality and dialogue-based systems, very sophisticated simulations can be expected to emerge that immerse the user in a convincing environment (Rickel and Johnson, 1998, Rich et al., 2002).

On the other hand, care still needs to be taken to ensure that the pursuit of realism does not overtake the learning goals. Increased fidelity and realism may increase user engagement. However, it may also increase the cognitive load. Under these circumstances, the core issues, the ability to reflect on these issues, and their relationship to the patient's well-being can become obscured. Therefore, appropriate evaluation still needs to be undertaken to make sure reflection is supported and taking place, and that learning is maximised.