

**The Design, Development and Evaluation of a Web-based Tool for
Helping Veterinary Students Learn How to Classify Clinical
Laboratory Data**

Jared A. Danielson

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Teaching and Learning

John K. Burton, Chair
Holly S. Bender
Katherine S. Cennamo
Glen A. Holmes
Barbara B. Locke
Gregory P. Sherman

July 7, 1999
Blacksburg, Virginia

Keywords: Web based Instruction, Clinical Pathology, Expert Scheme, Distance
Education, Applied Learning Theory

Copyright 1999, Jared A. Danielson

The Design, Development and Evaluation of a Web-based Tool for Helping
Veterinary Students Learn How to Classify Clinical Laboratory Data

Jared A. Danielson

(ABSTRACT)

Veterinary students face the difficult task of learning to classify clinical laboratory data. In an effort to make this task easier, a computer and web based tool known as the Problem List Generator (PLG) was designed based on current literature dealing with learning theory and medical education which are reviewed in chapter 1. Chapter 2 describes the design and the development process for the PLG. The PLG allows the students to access any number of cases (determined by the professor) of increasing complexity which provide signalment, history, physical exam, and laboratory data for a number of patients. Using the PLG, students analyze the data, identify data abnormalities and mechanisms, arrange them in a problem list, diagnose the problem, and compare their problem list and diagnosis to an expert problem list and diagnosis. The PLG was evaluated using a four step evaluation process involving an expert review, one-to-one evaluations, small group evaluations, and a two-part field trial, and was evaluated in terms of clarity, feasibility, and impact. The PLG is usable, in terms of clarity and feasibility, though fixes are recommended. There is no evidence to infer, statistically, that the PLG has any effect on learning outcomes. However, trends in the quantitative data and logical inference based on the context of the evaluation suggest that the PLG might help students, particularly those of low and average ability to produce more accurate problem lists.

DEDICATION

I dedicate this dissertation to my family, past, present and future. Thank you. And especially to my wonderful wife Britany. It's been an exciting ride so far. May the next several hundred years together be even better than the last few have been.

AUTHOR'S ACKNOWLEDGEMENTS

I gratefully acknowledge my advisor, John Burton. For me he has been an advisor, a leader, an example, a collaborator, and a friend. Thank you.

And to my other wonderful committee members from Teaching and Learning, Drs. Cennamo, Holmes, Lockee, and Sherman. Thank you for your time, encouragement, and confidence. You have helped make this an enjoyable and productive experience.

A special thanks to Dr. Holly Bender, who learned, with me, what it's like to see a dissertation to fruition in the Department of Teaching and Learning. Thank you for your boundless energy, enthusiasm, persistence, support and good humor.

I wish to acknowledge my parents. My father taught me how to write, for which I will always be grateful. Together they taught me to persevere, and to smile whether circumstances warranted it or not.

I wish to acknowledge Britany, Shayla, Adam, and Joseph, who with me paid the price for this and who deserve equal shares in the credit.

Finally, I acknowledge my God, upon whom, in the words of the ancient prophet Benjamin, I depend upon for my life, and for all that I have and am.

TABLE OF CONTENTS

DEDICATION	III
AUTHOR'S ACKNOWLEDGEMENTS	IV
LIST OF FIGURES	IX
LIST OF TABLES	X
CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW	1
THE INSTRUCTIONAL PROBLEM AND CONTEXT	1
REVIEW OF THE RESEARCH	1
<i>Learning Theory</i>	2
Behaviorism.....	2
Cognitive Perspectives.....	4
Constructivism.....	6
Theories involving Individual Differences	8
<i>Moving from Theory to Practice: Strategies and Models</i>	9
Strategies and Models Emphasizing Context.....	9
Situated cognition.....	9
Cognitive apprenticeships.....	10
Implications for the current problem.....	11
Anchored instruction.....	11
Implications for the current problem.....	12
Tessmer and Richey's General Model for Contextual Analysis for Instructional Design.....	12
Implications for the current problem.....	12
Landamatics and the Algo-heuristic Theory	13
Implications for the current problem.....	13
Models of Instructional Design.....	13
Instructional design models.....	14
Two examples of how to apply ID to a problem such as the	14
development of clinical problem solving skills	14
Implications for the current problem.....	16
<i>Synthesizing Theory, Models and Strategies</i>	16
"Active" Learning.....	17
Front-end Analysis	18
The Development of Expertise Takes Time.....	18
Context Matters	18
Evaluation.....	18
Summary of Synthesis with Implications for this Context.....	19
<i>Looking at the Literature</i>	19
Problem-Based Learning	19
A history and definition of problem-based learning.....	19
Has problem-based learning been successful?.....	20
How literature on PBL can help us understand the issues of problem classification.....	21
Knowledge Structures in Clinical Practice and Education.....	23
Learner models of reasoning as they relate to curricula.....	24
General models of diagnostic reasoning.....	25
<i>Synthesizing: Theory and Research</i>	27
CHAPTER 2: PRODUCT DEVELOPMENT.....	30

OVERVIEW OF THE PLG.....	31
<i>Step One</i>	32
<i>Step Two</i>	34
<i>Steps Three and Four</i>	35
<i>Step Five</i>	36
<i>Steps Six and Seven</i>	38
<i>Step Eight</i>	40
<i>Summary of the PLG</i>	40
DESIGNING THE PLG.....	40
<i>Interface Design</i>	41
<i>Instructional Design Models Employed</i>	41
THE DEVELOPMENT PROCESS.....	42
INITIAL EVALUATION EFFORTS AND MODIFICATIONS.....	42
<i>The Expert Review Process</i>	42
<i>The paper prototype</i>	42
<i>Evaluation of the paper prototype</i>	43
Participants.....	43
Observation.....	43
<i>One-to-one Evaluations of the PLG</i>	49
Participants.....	49
Observation.....	49
Clarity.....	50
Feasibility.....	50
Impact.....	50
CHAPTER 3 : EVALUATION METHODS AND RESULTS.....	51
SMALL GROUP EVALUATION - METHODS.....	51
<i>Participants</i>	51
<i>Procedures</i>	51
<i>Data Sources</i>	52
Introductory Note.....	52
Clarity.....	53
The first two small group sessions.....	53
The last two small group sessions.....	53
Feasibility.....	53
The first two small group sessions.....	53
The last two small group sessions.....	54
Impact.....	54
The first two small group sessions.....	54
The last two small group sessions.....	54
SMALL GROUP EVALUATION -- RESULTS.....	54
<i>General Results</i>	54
<i>Results from the Analysis of the Questionnaire</i>	55
<i>Results from observation, student comments, and the focus group conversations</i>	66
What they liked.....	66
What they didn't like.....	66
Items of disagreement.....	66
General reasons the students gave for dislikes.....	67
FIELD TRIAL - METHODS.....	67
<i>Introductory Note</i>	67
<i>Participants</i>	67
Group A.....	68
Group B.....	68
Group C.....	68
<i>Procedures</i>	68
Procedures for Students in Groups A and B.....	68

Procedures for students in Group C	69
<i>Data Sources</i>	69
Clarity	69
Feasibility	70
Impact	70
FIELD TRIAL EVALUATION -- RESULTS	71
<i>General Results</i>	71
<i>Specific Findings</i>	71
Impact	72
Clarity and Feasibility	75
Factors that lowered clarity and feasibility ratings	80
Factors that improved the clarity and feasibility ratings	82
Time Data as an Indicator of Feasibility	83
CHAPTER 4: DISCUSSION	84
IMPACT	84
<i>The Problem Lists</i>	84
<i>The Questionnaire</i>	85
CLARITY AND FEASIBILITY	85
CONCLUSION AND SUGGESTIONS FOR FURTHER EVALUATION	86
WORKS CITED	88
APPENDIX A -- INTERFACE DESIGN	93
INTRODUCTORY SUMMARY	93
DESCRIPTION OF HOW THE SYSTEM WORKS	93
<i>Overall Description</i>	93
<i>Data Classification and Synthesis</i>	98
<i>Evaluation</i>	104
<i>Other Functions</i>	106
Help	106
Print	107
Search	108
View my history	110
APPENDIX B - INSTRUCTIONAL DESIGN DOCUMENT	111
INSTRUCTIONAL DESIGN	111
<i>Overview of the Problem List Generator and Instructional Problem</i>	111
<i>Learner Analysis</i>	111
<i>Contextual Analysis</i>	111
Introduction to Context	111
Orienting Context	112
Learner factors	112
Immediate environment factors	112
Organizational factors	113
Instructional Context	113
Learner factors	113
Immediate factors	114
Organizational factors	114
Transfer Context	115
Learner factors	115
Immediate environment factors	115
Organizational factors	115
<i>Decision to Use a Web-Based Computer-Mediated Solution (Media Selection)</i>	115
<i>Instructional Goal</i>	115
<i>Instructional Analysis (See Figure B.1: Instructional Analysis)</i>	116

<i>Performance Objectives</i>	116
<i>Instructional Strategy</i>	118
Literature Review	118
Step-by-step Explanation of the Instructional Strategy.....	118
Summary of the Instructional Strategy	125
<i>Assessment Plan</i>	126
EVALUATION PLAN.....	127
1. <i>Expert Review</i>	127
2. <i>One-to-one Evaluations</i>	128
Participants	128
Observation	128
Clarity.....	128
Feasibility.....	134
Impact.....	134
3. <i>Small Group Evaluation</i>	134
4. <i>Field Trial</i>	134
5. <i>General Data Collection and Analysis for Estimating Impact of the PLG</i>	135
Reporting.....	136
APPENDIX B.1: SPECIFIC BENCHMARK TASK DESCRIPTIONS.....	138
<i>Benchmark Task #1</i>	138
<i>Benchmark Task #2</i>	138
<i>Benchmark Task #3</i>	139
<i>Benchmark Task #4</i>	139
<i>Benchmark Task #5</i>	140
<i>Benchmark Task #6</i>	141
<i>Benchmark Task #7</i>	141
<i>Benchmark Task #9</i>	143
<i>Benchmark Task #10</i>	143
<i>Benchmark Task #11</i>	144
APPENDIX C -- INFORMED CONSENT FORM.....	145
APPENDIX D: DIRECTIONS SHEET	147
APPENDIX E -- 1ST SMALL GROUP QUESTIONNAIRE	148
APPENDIX F: SECOND SMALL GROUP AND FIELD TRIAL QUESTIONNAIRE.....	150
USABILITY QUESTIONNAIRE (ACCOMPANYING USABILITY SPECIFICATIONS):	150
WORKS CITED -- APPENDICES A THROUGH F	153

LIST OF FIGURES

<u>Figure 2.1.</u> Case summary.....	33
<u>Figure 2.7.</u> Evaluation dialogue.	40
<u>Figure A.1.</u> Introductory Screen.....	94
<u>Figure A.2.</u> Initial Screen After Login	95
<u>Figure A.3.</u> “Open” Screen	96
<u>Figure A.4.</u> Case Summary	97
<u>Figure A.5.</u> Adding an item to the problem list from the lab data screen.....	98
<u>Figure A.6.</u> Lab Data.....	99
<u>Figure A.7.</u> The User’s Problem List.....	101
<u>Figure A.8.</u> Initial Problem List Generation	102
<u>Figure A.9.</u> My Problem List with Two Levels of Nesting	103
<u>Figure A.10.</u> “My Problem List” and “Expert Problem List” compared	105
<u>Figure A.11.</u> Evaluation Box.....	106
<u>Figure A.12.</u> Help Index.....	107
<u>Figure A.13.</u> Printer Interface	108
<u>Figure A.15.</u> Found Window.....	110
<u>Figure B.1.</u> Instructional Analysis.....	117
<u>Figure B.2.</u> Case Summary	119
<u>Figure B.3.</u> Adding an item to the problem list from the lab data screen	120
<u>Figure B.4.</u> Lab Data.....	121
<u>Figure B.5.</u> The User’s Problem List	123
<u>Figure B.6.</u> Completed Problem List	124
<u>Figure B.7.</u> “My Problem List” and “Expert Problem List” compared	125
<u>Figure B.8.</u> Evaluation Dialogue.....	126
<u>Figure B.9.</u> Usability Specifications	130
<u>Figure B.10.</u> Critical Incident Evaluation Sheet	133
<u>Figure B.12.</u> Sample Quiz.....	137

LIST OF TABLES

Table 2.1 <u>Usability Specifications: Evaluation Results -- 3 User Mean -- Paper Prototype</u>	45
Table 2.2 <u>Data Collection Sheet</u>	48
Table 3.1 <u>Means, standard deviation, median, maximum, minimum and construct assignment by item for Small Group Evaluation sessions 1 and 2 (From the questionnaire in Appendix E)</u>	56
Table 3.2 <u>Clarity, Feasibility, and Impact construct means for Small Group Evaluation sessions 1 and 2 (From the questionnaire in Appendix E)</u>	60
Table 3.3 <u>Means, standard deviation, median, maximum, minimum and construct assignment by item for Small Group Evaluation sessions 3 and 4 (From the questionnaire in Appendix F)</u>	61
Table 3.4 <u>Small Group Evaluation sessions 3 and 4 (From the questionnaire in Appendix F)</u>	65
Table 3.5 <u>Mean of all small group sessions</u>	65
Table 3.6 <u>Overall Means for PLG and Paper and Means for PLG and Paper by Ability Level Adjusted for Difficulty of Case</u>	73
Table 3.7 <u>Analysis of Variance for use of PLG Adjusted for Difficulty of Case</u>	74
Table 3.8 <u>Means, standard deviation, median, maximum, minimum and construct assignment by item for the Field Trial Evaluation Sessions (From the questionnaire in Appendix F)</u>	76
Table 3.9 <u>Questionnaire results for the students participating in the field trial for the constructs of clarity, feasibility, and impact</u>	80
Table 3.10 <u>Time Means Compared for the Field Trial</u>	83
Table 3.11 <u>Paired Samples T-test -- Time Taken to Complete the Problem Lists</u>	83

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

This dissertation is the result of an effort of one and one half years to address a real instructional problem in an authentic setting. In the pages that follow I will document the process and the results of this effort. This chapter begins with a discussion of the context and instructional/administrative problem that led to this instructional development effort. It then describes the research explored to help address the problem, and finally, presents a set of recommendations for resolving such a problem, based on the research.

The Instructional Problem and Context

Veterinary students at the Virginia Maryland Regional College of Veterinary Medicine (VMRCVM) face a new challenge the second semester of their sophomore year in vet school. In VM 8414 (Clinical Pathology) students are required for the first time in their veterinary classroom experience to synthesize the knowledge they have gained about anatomy, physiology, symptoms of medical problems, medical conditions, and characteristics of a variety of animals (species or breed-specific), in interpreting lab data and arriving at diagnoses. In an educational context which has favored rote memorization, the students frequently find this process difficult and frustrating. To help address this learning problem, a computer-based practice tool known as the Problem List Generator (PLG) has been designed and implemented.

The PLG also addresses an administrative problem. The professor wants each student to complete and receive credit for completing a problem list for between 6-9 cases per week. There has never been a feasible system for verifying that students have completed their problem lists, or for assigning credit for cases completed. The PLG allows students to do the problems on their own, evaluate their own responses, and record the fact that they have completed each assignment.

Review of the Research

The purpose of this chapter is to use relevant literature to guide the design of the problem list generator. Several related bodies of research are useful in approaching the problem of building a tool like the problem list generator. First, I will explore how several prominent learning theories deal with the issue of helping learners develop expertise in a complicated area such as clinical data classification. Second, I will look at general approaches, (strategies and models) used for dealing with such learning problems. Finally I will examine two bodies of research in detail that relate to learning in the domain of clinical problem solving. The first of these, broadly known as problem-based learning, deals with a certain approach to teaching medicine. The second related area of research attempts to identify the cognitive structures of medical practitioners who are successful and experienced.

Learning Theory

The theory of knowledge is full of paradoxes and conundrums, unsolved problems and difficult questions. Yet I think that there is a solid core of facts, important facts relevant to historically important questions, that are now beyond serious controversy. I think that we now know roughly how much we know and, rather more roughly, why we know it. My intention was to write down this solid core, and separate it out from the unsettled dubious stuff that usually surrounds and supports it. I found that much harder than I had expected. (Morton, 1977)

Anyone spending much time with current literature in fields related to theories of knowledge would sympathize with the difficulty Morton encountered in finding the solid core of facts "that are now beyond serious controversy." And yet, those in the field of teaching and learning must have at least enough solid notions about knowledge and how one acquires it to design and implement effective solutions to learning problems such as the one outlined at the beginning of this chapter. One of the primary difficulties inherent in obtaining consensus about how best to facilitate learning, is that it is currently impossible to explain learning in an uncontroversial way. This section provides a brief overview of some common ways in which modern theorists have attempted to explain learning, and some of the basic assumptions underlying those theories. I will frame my treatment of these issues within the context at hand. My purpose is not to provide thorough explanations of the theories, nor to even "cover" possible variations of each theory. Books have been dedicated to such endeavors. Rather, I will briefly examine several broad theoretical ideas, identifying how each would be relevant to teaching veterinary students how to classify clinical laboratory data. The categories that follow include behaviorism, cognitive psychology, and constructivism, followed by a section covering several theories/paradigms dealing with individual learner differences which, while not general theories of learning, are frequently used and discussed in education.

Behaviorism

One of the great dividing questions in a consideration of learning is to what extent it is useful to consider that which is going on inside the "black box" of the mind. This question frequently is considered another way: When examining learning, is it useful to consider anything other than learner behavior (i.e., learners' interaction with their environments)? In general, behaviorists would argue that the answer to this question is no (Amsel & Rashotte, 1984; Estes, 1950; Skinner, 1953; Skinner, 1954; Thorndike, 1971). Perhaps the following from Estes, (1950 p. 94) best illustrates this assumption: ". . . propositions concerning psychological events are verifiable only to the extent that they are reducible to predictions of behavior under specified environmental conditions. . ." In other words, if we are to understand "psychological

events" (i. e., learning), we had best stick to observing and predicting behaviors in controllable environments.

The work of B.F. Skinner can be used to illustrate this point of view. His theory, operant conditioning, has been widely applied in instructional settings (Kearsley, 1998). Skinner (1953) found that through carefully planned and administered "reinforcers" he could condition rats, pigeons, dogs, monkeys, and humans to perform any number of behaviors, some of which were very complex. He argued that changing behavior and learning are essentially synonymous. Therefore methods that are best suited, and most proven to change behavior in a predictable way are the methods most likely to be effective in teaching. He proposed a "mechanized instruction" system of education in which students would learn by interacting with machines which would be designed to manage the contingencies and reinforcers involved in any given learning task (Skinner, 1954). These would not replace teachers, but would leave them free "to function, not in lieu of a cheap machine, but through intellectual, cultural, and emotional contacts of that distinctive sort which testify to [their] status as a human being." (Skinner, 1954 p. 96-97).

Burton, Moore, and Magliaro (1996) indicate three assumptions of behaviorism that are directly relevant to designing instruction. **First**, the learner's role is active; "The learner must be engaged in the behavior in order to learn and to validate that learning has occurred" (p. 49). **Second**, "learning is a function of building associations between the occasion on which the behavior occurs (stimulus events) and the behavior itself (response events)" (p. 49). Complex learning (*such as the building of expertise in data classification*) involves a building of simple behaviors, the end result of which is the final goal. The contingencies associated with performing a certain behavior are not the same for all learners. Therefore, instruction must be tailored (designed) to fit any given learner, based on that learner's "genetic and reinforcement histories" (p. 49). **Third**, "the basic processes that promote or inhibit learning are universal for all organisms" (p. 50). This claim is supported by a substantial body of research showing a variety of organisms achieving equivalent learning results through equivalent methods (Skinner, 1954).

The mark of behaviorism can be found on most prominent models of instructional design which emphasize thorough and complete front-end analysis procedures (learner, content, task, and instructional analyses), the identification of specific behavioral objectives with related strategies, and a feedback mechanism which allows for altering the instruction until specific learning objectives are met (Gustafson & Powell, 1991).

To tie these ideas directly to the problem at hand, from the context of behaviorism, for a teacher/instructional designer to provide training which will cause learners to become expert at a complex task, they must do the following: a) Perform a solid front-end analysis, to understand all of the behaviors and sub-behaviors

involved in performing the complex task, as well as the characteristics of the "target" learners, b) create an instructional situation that requires the learners to practice/perform the appropriate behaviors, in proper sequence and with appropriate reinforcement, gradually building more and more behaviors until the target behavior is achieved, and c) verify that the target behavior has been achieved through evaluation and feedback. This process would require a great deal of time for complex, intricate tasks such as data classification.

Cognitive Perspectives

Cognitive psychology emerged in response to what many perceived as the limitations of behavioral psychology. Specifically, many felt that behaviorism was inadequate for explaining feelings, mental imagery, introspection and other phenomenon which could not be neatly tied to specific stimuli and responses, and, in some cases, seemed entirely unrelated to behavior (Winn & Snyder, 1996). These criticisms of behaviorism are by no means universally accepted, and the value of attempting to explain thought independently of behavior is still questioned by many (Burton et al., 1996). Nonetheless, cognitive psychology has gained an impressive following.

George Miller (1956) provided some early important research in cognitive psychology by introducing the concept of "the magical number seven" (p. 81), and the related concept of chunking. Essentially, Miller conducted a great number of studies which appeared to indicate that humans are only capable of remembering seven discrete bits of information without "processing" them somehow. Similarly, most peoples' span of judgement (their ability to discriminate between different classes of one type of data) is limited to seven. Miller concluded that people overcome this shortcoming by "chunking" information -- that is, recoding small bits of information into larger chunks which contain the smaller, related bits of information. The mind continues to chunk related chunks until all necessary information in one particular domain of knowledge is available to it at once. The resulting notion that the mind does something with information, organizationally, has intrigued many scholars and at present mental representations or metaphors to explain learning abound -- most of which rely heavily upon the concept of schema (Lindsay & Norman, 1977; Winn & Snyder, 1996).

A schema is a knowledge structure which accommodates/facilitates mental process. According to Winn and Snyder (1996) all descriptions of schema include the following: "(1) It is an organized structure that exists in memory and, in aggregate with all other schemata, contains the sum of our knowledge of the world. (2) It exists at a higher level of generality, or abstraction, than our immediate experience with the world. (3) It consists of concepts that are linked together in propositions. (4) It is dynamic, amenable to change by general experience or through instruction. [and] (5)

It provides a context for interpreting new knowledge as well as a structure to hold it" (p. 117).

An explanation of mental structures and processes, then, is the stuff of cognitive psychology, and such a study, given the difficulty inherent in studying and explaining that which is unobservable, cuts a rather wide swath. Naturally there are many flavors of cognitive theory. Some researchers have attempted to build general models of cognition that explain or simulate mental processes. Anderson's ACT model, Newell and Simon's GPS (General Problem Solving) model, R. Schank's Script theory, and Paivio's Dual Coding theory are all metaphors which attempt to explain, in a variety of ways how the mind works (Clark & Paivio, 1991; Kearsley, 1998). Some of these theories, like Dual Coding theory, are meant to provide a relatively complete model of cognition (Clark & Paivio, 1991). Others, like the GPS model, are meant to model specific mental phenomena (such as problem solving, in this case) (Kearsley, 1998).

Given the eclectic nature of the models and concepts in cognitive psychology, it is difficult to summarize how the concepts involved might be applicable to a problem involving teaching the difficult process of data classification and synthesis to novice learners. In essence, however, the alteration of a schema is a process that takes time. And the development of expertise is dependent upon the alteration of schema. Just as Piaget proposed that all humans go through stages of development as their schemata mature and grow (Lindsay & Norman, 1977), the schemata of learners are molded and shaped in stages as they develop expertise (Winn & Snyder, 1996). Norman (1982) categorizes this process into three "modes": accretion, structuring, and tuning. Accretion involves adding information to existing schemata, structuring involves the formation of new schemata, and tuning involves "fine adjusting of knowledge to a task" (p. 81). Tuning is seen as a time-consuming process, generally involving hours of practice in which general knowledge and schema are appropriate to the task, but the knowledge is not a perfect fit to the task, and so must be adjusted accordingly.

From Winn and Snyder's (1996) perspective, this understanding of the nature of expertise in relation to mental structures implies changing behaviorally-based instructional design (as they perceive it) in several ways. First, they propose that educational technology, until "recently" has ignored the potential of technology to help people learn anything other than "objective facts and rules." They suggest that such training has essentially entailed "creating nothing more than training wheels for learning and acting that are to be jettisoned for more important knowledge and skills as quickly as possible" (p.124). The authors do not provide specifics, making it a little difficult to imagine how one would go about implementing changes to instructional design for teaching knowledge and skills that are simply described as "more important." Nor does it seem likely that many informed scholars in the field of

instructional technology would agree that most instructional design until recently was limited to the teaching of "objective facts and rules."

Winn and Snyder's (1996) second "implication" for instructional design has a few more teeth. In general, they propose that, since the road from novice to expert is a multi-stage process, training cannot be based on expert skills, but must reflect the entire process through which the learner must pass. While they do not spell it out in so many words, the implication is that instructional designers must understand and, somehow design instruction for, the entire process that leads from being a novice to being an expert. This in turn implies that an instructional or task analysis based completely on expert performance is inadequate. In Winn and Snyder's words, "If we try to teach the skills of the expert directly to novices, we shall surely fail" (p. 125).

In a later section we will explore a specific approach that one cognitive psychologist, Lev Landa, has taken to deal with the process of helping learners gain expertise in dealing with relatively complicated processes.

Constructivism

The movement which brought us cognitive psychology as discussed above was, in the words of Jerome Bruner (1990), "intended to bring the 'mind' back into the human sciences. . ." (p. 1). And while not critical of the net gain in knowledge that he feels emerged from this movement (which he calls the "cognitive revolution"), Bruner asserts that the cognitive revolution did not achieve its original purpose. That purpose? "To establish meaning as the central concept of psychology" (1990, p. 2). And the construction of meaning, to Bruner, is inextricably tied to the individual's intentional states which are formed "only through participation in the symbolic systems of the culture" (1990, p. 33). Similar thinking originates with Vygotsky, who asserts that "mental functioning occurs first between people in social interaction and then within the child on the psychological plane", and that "the very structure of individual functioning derives from and reflects the structure of social functioning" (p. 1-2) (Wertsch & Rogoff, 1984).

To the constructivist then, there is no point in studying learning solely as behavior or solely as information processing. Because, to the constructivist, meaning, thought, and individual functioning cannot be explained independent of intent and culture.

Applications for and explanations of constructivism abound. Constructivism is viewed and applied in various ways including, but not limited to, the following: a) as a paradigm for research (Guba & Lincoln, 1994), b) as a cognitive science (Bednar, Cunningham, Duffy, & Perry, 1995), c) as a slogan meant to imply "caring" (particularly in educational circles), (Duffy & Cunningham, 1996), and d) as a theoretical framework for designing/facilitating instruction (Duffy & Cunningham,

1996). It borrows from psychology, anthropology, sociology, philosophy, and "the humanities" (Bruner, 1990).

While one might sooner hope to strike a galloping antelope with a peashooter as arrive at consensus for one theory of constructivist instruction (the term "constructivist instruction" itself would be seen as a contradiction in terms by some), I will report here some implications of constructivism for instructional design as proposed by some leading constructivist scholars in the field of educational technology.

Duffy and Cunningham (1996), present a "mind as rhizome" (MAR) metaphor for the mind, and offer it as a constructivist metaphor. I cannot do justice to a complete description of the MAR here. But, in summary, the MAR metaphor presents the mind as a constantly shifting structure without fixed points or hierarchy, any point of which can at any time interact with anything with which it comes in contact, both within and outside of the mind itself. Thinking occurs at the points of interaction.

To Duffy and Cunningham, the model for instruction that is most consistent with their MAR metaphor of the mind is what they call "The Problem as a stimulus for authentic activity" (1996, p. 190). In this model, instead of "teaching", the teacher/facilitator supports the student's learning "as skills are developed through working on the problem" which the authors equate to "authentic activity." They argue that "discovery learning", "scaffolding", "cognitive apprenticeship", "coaching" and "collaborative learning", all frequently referred to as "constructivist" methods, are only truly constructivist to the extent that they are used not to teach students what they "should do/know and when they should do/know it" (p. 191), but rather to "support the students in developing their critical thinking skills, self-directed learning skills, and content knowledge in relation to the problem" (p. 191). Their problem-based learning (PBL) model is based heavily on Barrow's problem-based learning curriculum as implemented in medical education. (Research on medical problem-based learning will be discussed in some detail later in this paper.) Duffy and Cunningham (1996) identify the key issues in designing PBL instruction as: a) task analysis (identification of key concepts and their professional use) b) problem generation (problems must be real and deal with relevant concepts and principles) and planning carefully for c) the learning sequence, d) the facilitator role, and e) assessment.

The above model is presented as one with general applicability, and so could be seen as a constructivist solution to the particular learning problem which provides the framework for this paper -- that of helping veterinary students deal effectively and accurately with clinical lab data.

Theories Involving Individual Differences

The theories discussed up to this point are all based on the assumption that there are certain principles of learning that are common to all learners. Here I will briefly mention several models related to learning that focus on individual learner differences.

Aptitude-treatment interaction (ATI) as presented by Cronbach and Snow (1981) is a framework for conducting research based on the assumption that fundamental learner differences will affect what instructional methods are most appropriate. A variety of experimental designs are put forth for analyzing learner differences, but the authors caution that ". . . one can be pessimistic about lofty aims for 'new paradigms' or 'integrations'" based on ATI research because "the complexity of demonstrable interactions is sufficient to exceed the practical limits of experiments designed to unravel them" (p. vii). The task of "unraveling" learner differences and how they are affected by differing instructional "treatments" continues to be a thorny (and political) one, and the possibility of a research-based theory of learning that satisfactorily addresses even common/generic learner differences such as age, gender, and "learning styles" (another problematic term) seems distant at best.

Another area of teaching and learning that focuses on learning differences is adult learning. There do not appear to be any universally accepted "theories" of adult learning. Knowles' (1984) theory of andragogy has been offered as such, though Knowles himself acknowledges that the question of whether or not andragogy is a theory is a controversial one, and himself claims not to know whether it is a theory or not. In general, scholars in adult learning propose that there are certain characteristics of adult learners that call for specific instructional processes. Regardless of their "theory" status, the assumptions of Knowles' model are fairly representative of adult learning as a whole (Brockett & Hiemstra, 1991; Cross, 1981), and I will briefly present them here. These assumptions are: a) the learner is self-directing, b) adult learners have more and different experience than youth, c) adults are ready to learn when they need to know to perform better in their lives, d) the adult orientation to learning is "life-centered", "task-centered", or "problem-centered", and e) adults tend to be internally motivated (Knowles, 1984).

Adult learning and ATI are only two of several perspectives on learning that focus on individual differences. Gardner's Theory of Multiple Intelligences, Sternberg's Triarchic theory, and Guilford's Structure of Intellect theory all focus on various ways of viewing and dealing with individual differences, though of these three, Gardner seems the most willing to describe how his theory would affect instruction. Sternberg and Guilford's theories seem more focused on explaining individual differences than in addressing their implications for instruction (Kearsley, 1998).

The individual differences theories mentioned above have potential implications for teaching veterinary students how to classify clinical lab data. For example, veterinary students are adult learners, and many have come to school with a very real set of problems that they hope education will solve. It is possible that some of the literature related to adult learning will be relevant here. It should also be noted that instructional design models based on behaviorism, cognitive psychology, and constructivism, all call for taking measures to account for learner characteristics.

Moving from Theory to Practice: Strategies and Models

Discussing theory without discussing at least general principles of practice is a difficult proposition. For example, though I began the previous section with the intention of discussing theory alone, general ideas of how each theory might inform practice seemed to naturally follow the discussion of each. This section is dedicated to a more in-depth exploration of how several theoreticians and practitioners in the field of teaching and learning have proposed to address problems similar to mine. I will then move to an exploration of specific relevant research in the domain of medical education.

Strategies and Models Emphasizing Context

The importance of allowing for context in instructional design has long been recognized, and is appreciated by behaviorists, cognitivists, and constructivists alike. However, the importance of planning for and including social, cultural and task-related contextual factors in instruction is receiving heightened attention in recent years, and strategies which focus on them tend to be tied to constructivism. Several sets of ideas related to the role of context in instruction seem worthy of note here. First I will discuss the general idea of situated cognition, followed by two examples of strategies that attempt to employ it: cognitive apprenticeship and anchored instruction. I will then discuss Tessmer and Richey's (1997) General Model for Contextual Analysis for Instructional Design, a more traditional and thorough approach to allowing for context .

Situated cognition.

Situated Cognition, in and of itself, cannot really be considered a strategy for instruction. Rather, as explained by the Cognition and Technology Group at Vanderbilt (1990) situated cognition "provides a broad, useful framework that emphasizes the importance of focusing on everyday cognition, authentic tasks, and the value of in-context apprenticeship training" (p. 2). The purpose then, of strategies based on situated cognition, is to provide, in a formal learning setting, the kind of authentic, "everyday" learning experience that might be found in actually doing what is being learned in a real-life context. Brown, Collins, and Duguid (1989) introduce

this concept as "enculturation" and argue that to become expert in a particular skill, one must become a part of the culture of those who are experts. Situated cognition is popularly associated with a constructivist philosophy, though the basic idea that a learning environment should closely approximate the "real" setting, or that learning is best accomplished in "real" settings could be found in behaviorist or cognitive literature. One essential difference between "real" settings to the constructivist and the behaviorist, is that the behaviorist feels at liberty to create a learning environment in which contingencies are allowed for and appropriate stimuli are provided.

Two strategies commonly associated with situated cognition are cognitive apprenticeships and anchored instruction.

Cognitive apprenticeships.

Collins, Brown, and Newman (1989) summarize how effective cognitive apprenticeships occur under the broad category of "characteristics of ideal learning environments" with subcategories including "content", "methods", "sequence" and "sociology", each of which contain additional subcategories, which I will briefly summarize below:

Content:

Domain Knowledge: Conceptual and factual knowledge and procedures -- insufficient on its own for problem solving.

Heuristic Strategies: "Tricks of the trade" -- effective "shortcut" techniques and strategies.

Control Strategies: Control the process of problem-solving. Include diagnosis, monitoring and remediation.

Learning Strategies: Strategies for learning any of the above content.

Methods:

Modeling: The expert carries out the task -- learners observe.

Coaching: The experts observe students as they practice while giving hints and providing scaffolding, feedback, modeling, reminders, and new tasks (p. 481).

Scaffolding: Any teacher support for carrying out the task -- suggestions, help, cue cards, etc.

Articulation: Involves getting students to articulate their knowledge.

Reflection: Students compare their problem-solving processes with expert processes.

Exploration: Pushing students to do their own problem solving through setting goals and sub-goals.

Sequence:

Increasing complexity: Tasks and task environments move from simple to complex.

Increasing diversity: Tasks and task environments move from requiring few to more strategies and skills.

Global before local skills: Students apply skills in an interesting problem before they have to generate and remember the skills.

Sociology:

Situated Learning: Practice and learning environment reflects the multiple uses to which knowledge will be put in the future.

Culture of expert practice: Active communication and skills practice in the area of expertise. Expertise is seen as "the practice of solving problems and carrying out tasks in a domain" (p. 488).

Intrinsic motivation: Intrinsic motivation is promoted.

Exploiting cooperation: Atmosphere of cooperative problem solving.

Exploiting competition: Comparison and assessment of student processes in solving identical problems.

Presumably, good teaching/learning built on a model of cognitive apprenticeship will include some or all of the above elements. However, the above set of suggestions are apparently meant to provide a set of tools, rather than a prescriptive model, for the examples of good practice provided by the authors do not appear to employ all of the above strategies, and include some strategies not explicitly listed above. The examples of cognitive apprenticeship provided appear to rely on relatively small class sizes, and a great deal of interaction between teacher and learner. The authors also fail to deal with the common problem of the expert's frequent inability to articulate the cognitive strategies involved in expert performance (Speelman, 1998).

Implications for the current problem.

Of the three context-based approaches to instruction explored here, cognitive apprenticeship is the most directly related to the development of expertise. While the problem list generator is planned to be delivered via computer technology, and the authors do not deal specifically with computer assisted instruction, many of the principles of cognitive apprenticeship (such as scaffolding, building, modeling, exploration, and sequencing principles) could be, and have been employed in computer assisted instructional environments.

Anchored instruction.

Anchored instruction, introduced and championed by the Cognition and Technology Group at Vanderbilt (1990; 1993), might best be considered a strategy for teaching based on assumptions about the situated nature of learning. Anchored instruction is built on video-based anchors which provided a "macrocontext" for instruction. These "macrocontexts" consist of stories whose contexts are rich with

information and lend themselves well to problem generation and problem solving. Students are generally asked either to explore a context (story) generally, glean information through answering general questions about it, or to answer specific questions which require a good deal of exploring and problem solving to answer.

Implications for the current problem.

Anchored instruction has been applied with children in formal school settings, and the extent to which it has been used for training for specific tasks is unclear. The lack of a duplicable "model" for using anchored instruction, other than the example descriptions given by the authors makes it difficult to apply other than in a general sense.

Tessmer and Richey's General Model for Contextual Analysis for Instructional Design.

Tessmer and Richey (1997) assert that while traditional instructional design models have dealt with context, they do not adequately provide for designing instruction that allows for contextual factors. Traditional contextual analysis has limited itself largely to an examination of what the learners know coming into the training, general demographic information about the learners, and at times in-depth examinations of the environment in which the learning task will be performed once learned.

They argue that a more in-depth examination of context is needed, and that frequently, the failure of instruction is due to an incomplete analysis of context. To Tessmer and Richey (1997), there are three levels of context with which the designer must be concerned: the learner, the learner's immediate environment, and the organizational environment. Furthermore, each of these contextual levels must be explored and planned for in three different contexts: the orienting context (where the learner is coming from), the instructional context (where the learning is taking place) and the transfer context (where the new skill will be employed on a daily basis). For a complete contextual analysis, then, nine contexts (learner/orienting, learner/instructional, learner/transfer, immediate environment/orienting, immediate environment/instructional, immediate environment/transfer, organizational environment/orienting, organizational environment/instructional, and organizational environment/transfer) must be explored and planned for. In their model, Tessmer and Richey provide a detailed framework for conducting a contextual analysis and suggest contextual analysis tools (1997).

Implications for the current problem.

This model provides a useful way to put some of the rather fuzzy ideas about situated cognition into practice in a variety of instructional design settings having a

variety of contextual constraints. It would be useful to use the model, if for nothing else, to make sure that all aspects of context are planned for in building any given instructional tool, including the problem list generator.

Landamatics and the Algo-heuristic Theory

Landamatics refers to a method of analysis and instruction for improving novice performance (Technology, 1993). Landamatics is based on cognitive psychologist Lev Landa's training application of his algo-heuristic theory. The algo-heuristic theory asserts that mental functioning can be seen in terms of mental operations and knowledge units. All performance problems can be seen as algorithmic, semi-algorithmic, heuristic, or semi-heuristic (Kearsley, 1998). Many problems are solved using expert "algorithms", which can be determined through careful observation/questioning of the expert. Designers discover the algorithms experts use to do their work, and train novices in their use. After such training, novices can perform as experts (and, in cases where several expert algorithms have been combined to form a super-expert algorithm, better than experts) (Technology, 1993). Some problems are too complex or creative to be solved with algorithms. Such problems are dealt with using expert heuristics. Other problems involve both algorithms and heuristics. Landamatics has been applied mostly to problems of the algorithmic variety, with great success. In 1993, (Technology), Landa reported phenomenal success using Landamatics in a variety of industrial settings, including the establishment of the "Allstate (insurance company) Landamatics University".

Implications for the current problem.

The analysis of clinical laboratory data is an extremely complicated process. It is certain to contain many sets of algorithmic and heuristic problems. So, the extent to which a straight "Landamatics" approach to teaching such a skill would be feasible or desirable is not clear. However, there is no question that a complete exploration of the expert processes involved would aid any effort to teach such processes.

Models of Instructional Design

Creating one heading called "models of instructional design" and stuffing below it some of the most in-depth, tried and tested, and thorough work on solving educational problems available can only be called an injustice. And so, even before committing the act, I will present what I hope will be an adequate explanation for doing so. There are a number of similar instructional design models which borrow from similar assumptions of behavioral, cognitive, and systems theory (Gustafson & Powell, 1991). While all of these models make unique and important contributions to instructional design theory and practice, they are similar enough that a relatively detailed discussion of one or two will serve my purposes here. (Which are to outline how such models might contribute to the task at hand).

Conscience eased, I will carry on with a general overview of systemic models of instructional design, followed by examples of how two prominent thinkers in instructional design literature, Robert Gagné and David Merrill, might approach a problem such as developing expertise in clinical problem solving.

Instructional design models.

Dick and Carey (1996) define a system as "a set of interrelated parts, all of which work together toward a defined goal" (p. 3) A systems model for instruction is one in which a variety of designer activities work together for the hoped for result of effective instruction. These "designer activities" vary from model to model, depending on the specific situation in which the model is to be applied, and the specific assumptions about learning that underpin the model. However, at the heart of these models are several common processes, including some sort of front-end analysis to find out about contextual factors and goals, a process for creating materials or instruction, and a feedback/evaluation mechanism to determine if the process is working. The Dick and Carey Systems Approach Model for Designing Instruction (1996) is commonly used as a representative instructional design model, and for teaching instructional design to novice instructional designers. Dick and Carey's model includes nine related processes, which, under simple or uncomplicated circumstances, might proceed somewhat linearly. Those processes are: 1) determine instructional goal, 2) analyze the instructional goal, 3) analyze learners and contexts, 3) write performance objectives, 4) develop assessment instruments, 5) develop instructional strategy, 6) develop and select instruction, 7) design and conduct the formative evaluation of instruction, 8) revise instruction, and 9) conduct summative evaluation. It must be remembered, however, that, because the model is systemic, all related parts "work together" and the order and frequency of their performance is dictated by the particular context of the learning problem being addressed.

Dick and Carey promote their model as being appropriate for use by instructional designers, instructors, public school teachers, university professors, industrial trainers, and military instructors. Like most of the instructional design models I am using it to represent, it is designed to be applicable for a wide variety of learning tasks and in a variety of situations. There is no question that there is much in such a model that would be applicable to the creation of the problem list generator.

Two examples of how to apply ID to a problem such as the development of clinical problem solving skills

1. Robert Gagné

Suppose I were to turn to Gagné, Briggs and Wager's (1992) "Principles of Instructional Design" for solutions in teaching veterinary students how to classify clinical lab data. According to Gagné et al. (1992), there are five kinds of learned capacities: intellectual skills, cognitive strategies, verbal information, motor skill, and attitude. An examination of my problem would reveal that to successfully classify

clinical lab data, one must have a great deal of specific verbal information, the appropriate attitude, a variety of intellectual skills, including higher order rules (and, perhaps, problem solving) as well as, in some cases, cognitive strategies. Following the authors' framework, I would define performance objectives, analyze the overall learning task (in a more specific way than I've done already), design instructional sequences, select appropriate media, design the individual lessons, and assess the instruction.

Of course, the process I've just summed up would be a lengthy and circuitous one. To provide an example, I'll look at one of my trickiest goals. I've decided that I want to teach veterinary students how to "think" like experts. This particular goal involves cognitive strategies. Gagné et al. warn that the research on teaching cognitive strategies is not encouraging if what I'm after is a quick simulation that will teach all students how to (for example) synthesize and find connections between data regarding the functioning of any of a number of body systems and organs. First, they warn, cognitive strategies are, at least partially, a function of genetic factors. Second, "the design of instruction for cognitive strategies has to be done in terms of 'favorable conditions'" (p. 70). In other words, if I want to teach cognitive strategies to vet students, I'm going to have to give them opportunities and situations in which they are forced to practice the cognitive skills in question. (And suddenly I'm finding myself led by a "behavioral/cognitive" tradition down a road which could easily be called "situated cognition" from a constructivist tradition.)

2. David Merrill

In a recent conversation which I was fortunate enough to have with David Merrill, I briefly presented the problem of teaching how to classify clinical lab data to veterinary students, and asked how he would go about addressing it. He referred me, among other things to a book he wrote with Robert Tennyson in 1977: "Teaching Concepts: An Instructional Design Guide", and suggested I use it as part of the solution to my problem. It should be noted that Merrill is not a man who is easily daunted by the complexities of learning, and this is reflected in the way he approaches learning problems. (His current project involves building an instructional design expert system which would, among other things, sequence instructional transactions, interact with the learner, and adapt the way transactions are enacted to meet individual student needs (Merrill & Group, 1996).

In "Teaching Concepts . . ." Merrill and Tennyson (1977) define a concept as "a set of specific objects, symbols, or events which are grouped together on the basis of shared characteristics and which can be referenced by a particular name and symbol." In the context of clinical lab data, and using this definition of concept, any particular body organ or system, as well as the specific ways in which they interact, or fail to interact correctly, could be defined as concepts. Having **a**) determined that I am, indeed dealing with concepts, I would then proceed to **b**) identify, define and

name all the concepts I want to teach. For example, a rather complicated concept, might involve salmonellosis, which is associated with other "concepts" including hyponatremic dehydration (from diarrhea), an acute inflammatory leukogram and toxemia, etc. I could also identify all specific relationships between these concepts, which are, themselves definable and nameable concepts. In my case I would, conservatively, be dealing with thousands of concepts. Having defined the concepts, I would now **c**) collect an "instance pool" (large set of examples and non-examples) for each concept. For a simple example, hyperventilation is an example of a condition associated with the concept "alkalemia", while hypoventilation is a non-example. Having accumulated a variety of instances for each concept (thousands of total instances), I would now **d**) estimate the difficulty for each instance (the instances I've created should be of a variety of difficulties). I would now **e**) prepare a diagnostic classification test for each instance which "enables the instructor to make valid and reliable inferences about the student's ability to classify newly encountered instances of the concept." Merrill and Tennyson's model for teaching concepts continues in like manner, through careful definition of strategies for teaching, to identifying relationships between concepts, to stages of evaluation. The complexity of approaching my learning problem this way is dizzying when viewed in terms of teaching discrete concepts. And yet, an elephant is eaten one bite at a time, and there is no question that such an approach would be thorough and would result in positive, measurable learning outcomes.

A comparison of Gagné et al.'s prescriptive, yet principle-based approach to instructional design and Merrill and Tennyson's "no stone unturned" approach highlights some of the commonalties and differences among "traditional" approaches to instructional design, of which, as I stated at the beginning of this section, there are many.

Implications for the current problem.

There are a number of models of instructional design which provide generally prescriptive strategies and procedures which are likely to produce useful learning outcomes when applied to the problem of creating an on-line tool designed to teach veterinary students how to classify clinical laboratory data.

Synthesizing Theory, Models and Strategies

There are those who argue that instructional interventions are weakened or "stripped of meaning" if they borrow from a variety of theoretical perspectives (Bednar et al., 1995). It doesn't matter, they assert, what the particular theoretical base may be, but it must be kept pure in practice. I respect the importance of tying practice to a particular theory, and certainly thoughtless and theory-less instruction should prove less effective than instruction which is strongly tied to a strong theory. On the other hand, as I hope my brief treatment of learning theory has illustrated, we

are still far from a set of ideas about learning which are, if I might borrow Morton's words, "beyond serious controversy" (Morton, 1977). And, as pointed out by Skinner (1974) in the early days of the "cognitive revolution" and reiterated more recently by Burton et. al (1996 p. 65), early western medical procedures which included ". . . bloodlettings, leechings, cuppings, poultices, emetics, and punctions. . ." had solid theoretical moorings in the understanding of the human body of the day. That those treatments frequently resulted in death did not affect their theoretical validity. I think it can be argued that learning theories of today are more firmly rooted in research than were seventeenth century western medical interventions. However, though by so saying I risk broadcasting a naively positivist orientation, (Guba & Lincoln, 1994) we really do not know if any of our relatively new theories of learning are "right." And so, while not a "pure" approach, I would argue that the "safe money" is on an educational intervention that can be defended from a variety of theoretical perspectives.

Having planted myself philosophically, I will now bring together the theoretical perspectives presented in this paper, as they relate to the problem of teaching veterinary students how to classify clinical lab data. The theories do indeed come together with several important implications for instructional design practice.

"Active" Learning

Both behavioral and constructivist positions require an "active" learner. From the behavioral perspective, since learning occurs through/is manifested in behavior, the learner must be doing in order to learn and show learning. From the constructivist perspective, learning occurs best in authentic environments, in which learners are doing the thing they are learning how to do.

From a cognitivist perspective, of course, it is hard to imagine "passive" learning because the internal mental processes are always "active." The literature on implicit mental processes does suggest that, in some cases, learners learn without ever being aware of what they are learning (Speelman, 1998). In these cases, however, we are dealing with an unaware, active learner, not a passive learner. In other words the learner is actively performing tasks which result in an acquisition of knowledge or skills of which the learner is unaware, in spite of possessing them. This process, incidentally, is completely consistent with behavioral or constructivist learning theory.

Regardless of the theoretical perspective, then, it seems clear that the learner should be actively engaged in the task to be learned, or some sub-task of the task to be learned.

Front-end Analysis

All of the literature is in agreement as to the importance of establishing an understanding of expert performance. To the constructivist, this understanding is essential to create an appropriate authentic environment. To the cognitivist, a number of analyses are called for, so that the designer understands not just the expert schema but the schema-altering process that lead from the beginning knowledge base to expertise. To the behaviorist, a similar analysis is required -- seen in terms of tasks and skills and subtasks and subskills.

Regardless of the theoretical perspective of the designer, then, it would be important for him/her to conduct a detailed front-end analysis.

The Development of Expertise Takes Time

The fact that the development of expertise is a lengthy one, is fairly well established in research. In summarizing the research on the development of expertise in a complex area, Burton and Magliaro (1988) indicate that it generally takes between 5,000 and 10,000 hours of practice with feedback to achieve expertise. Behaviorism, cognitivism, and constructivism all easily accommodate this research, for whether one sees learning as the adoption of complex behaviors, the alteration of schema, or the socially negotiated construction of reality, it is easy to explain the process as a lengthy and complicated one for complex tasks.

Context Matters

From the behavioral position, of course, context (everything with which the learner interacts) is everything, though behaviorists feel at liberty to create an effective "artificial" instructional context. Constructivists value the importance of context as well, and have drawn attention to the societal aspects of context. Perhaps the cognitive position is the least dependent on context, though it seems unlikely that a cognitive scholar would disparage the usefulness of having "situated" instruction. So, while specific estimations of how much context matters or what context is and is not might vary from one theoretician to another, it seems unlikely that any would be found who disparage the usefulness of providing some degree of authenticity in the learning environment.

Evaluation

While learning can certainly occur without evaluation, most design efforts, regardless of the designer's assumptions, include formative and summative evaluation to verify the results of the teaching effort.

Summary of Synthesis with Implications for this Context

In summary, a theoretically effective instructional intervention for teaching veterinary students how to classify lab data will be based on a solid front-end analysis, will allow for context, will required learners to do real tasks in real/realistic settings, and will allow/require the necessary time for the learning goals to be achieved.

Looking at the Literature

Having established a theoretical case for the design of the problem list generator, I now turn to the literature examining the teaching of skills such as data classification in the field of medical education to determine to what extent the theory and the empirical literature are consistent. The specific bodies of literature which seem particularly relevant here are those dealing with problem-based learning and how it compares to a traditional medical education curriculum, and literature that deals with the acquisition of expertise in the field of medicine.

Problem-Based Learning

A history and definition of problem-based learning.

Problem-based learning (PBL), in the realm of medical education, began in the late 1960's at the Medical School of McMaster University in Canada. McMaster introduced PBL as a response to the conventional curriculum of the day. The conventional curriculum (still practiced in many colleges of human and veterinary medicine) tended to focus on the learning of isolated basic science facts followed by clinical experiences. Problem-based learning is based on the premise that medical education would be more effective if medical facts were learned, not in isolation, but in the context of clinical cases. Therefore, in a PBL setting, students learn and are self-taught in the context of medical "problems." Students collaborate, consult textbooks and other resources, and teach themselves (with professor guidance and direction) (Albanese & Mitchell, 1993; Berkson, 1993; Norman & Schmidt, 1992; Rivarola, Bergesse, Garcia, & Fernandez, 1997; Rosing, 1997; Wilkerson, Hafler, & Liu, 1991).

Arriving at a commonly agreed upon working definition of PBL is problematic. In the 1960's problem-based learning was, by definition, what was done at McMaster Medical School. While efforts have been made to document what is done at PBL schools, a perfectly replicable and exportable model of any curriculum is difficult, at best, to achieve. Furthermore, as any educational researcher knows, any given curriculum, even within the confines of one institution, is going to vary in practice from teacher to teacher and class to class. The difficulty of defining problem-based learning is further complicated by the fact that, by September of 1992, about

sixty Medical Schools in the world had embraced problem-based learning, each with its own unique approach to the method. So much have these approaches departed from the original, that at least one scholar was moved to create a taxonomy of types of problem-based learning (Albanese & Mitchell, 1993). A general definition of PBL by Vasconez, Donnelly, Mayo, & Schwartz (1993), which seems consistent with most PBL models, claims PBL to be “an attempt to simulate the experiences and thought processes that ‘real-life’ physicians go through when facing patients who present challenging problems. PBL places a small group of students in situations in which they must identify the major problems, learning issues, and solutions pertinent to a series of cases taken from real scenarios.” (p. S28). The professor, in a PBL environment “(1) actively participates in the design of multidisciplinary learning programs, which are usually composed of various faculty disciplines, (2) guides the students through the learning programs, and (3) hardly gives any lectures” (Rosing, 1997 p. 71).

The idea of having students deal with real-life problems in a self-directed fashion with minimal supervision in a small-group, collaborative environment is certainly not new to those who are familiar with teaching and learning, nor was it new in the late 1960’s. However, new to the world of medical instruction was the concept that students, if given situations in which they were required to act as experts, would be more likely to develop “expert” ways of solving medical problems. This was the hope and promise of problem-based learning.

Has problem-based learning been successful?

In spite of the fact that it continues to be adopted by leading medical schools such as the Harvard School of Medicine (Albanese & Mitchell, 1993; Berkson, 1993), PBL has not been shown to significantly enhance the performance of its students or graduates (Albanese & Mitchell, 1993; Berkson, 1993; Norman & Schmidt, 1992). This could be attributed to a number of factors, including the fact that PBL is so hard to define and is practiced differently at different schools of medicine. Another issue faced by those attempting to assess the usefulness of PBL is that most of the studies examining PBL have not had the design strength to support causal claims. One major difficulty faced by researchers of PBL is that of inherent population biases. Very few medical education programs run concurrent PBL and traditional curricula, and even in cases where this occurs, students are allowed to select which curriculum to participate in, and so studies comparing the two are fraught with the likelihood of built-in error. In a study of one such situation, (Michigan State, where students are offered a choice of a PBL or conventional curriculum), Bridgham, Solomon, and Haf (1991) found that students with significantly low entrance exam scores systematically selected themselves into the conventional instructional track. Therefore, even in fairly robust studies, researchers are limited by the fact that they must work with intact groups -- those whom have been recruited and accepted by a given medical school. Such groups are predetermined, and therefore, have varying systematic characteristics, such

as G.P.A., standardized test scores, and demographics. Random assignment of a population of study, use of control groups, etc. (and, therefore, the ability to imply causality) is nearly impossible in a study of PBL (Patel, Groen, & Norman, 1991).

Many researchers have turned to case studies in their attempts to explore PBL. Unfortunately, while case studies can provide a reasonably accurate picture of what is occurring, they are limited in their ability to generalize.

How literature on PBL can help us understand the issues of problem classification.

While research on PBL has some limitations, it is useful to one attempting to ferret out information necessary for building and evaluating a tool such as the problem list generator. One major question to answer is, “how do students develop expert data classification skills?” Because medical education is generally concerned with this kind of skill, and because PBL has been studied extensively, an analysis of research on PBL should be helpful in addressing this problem. In general, research on PBL has demonstrated several important concepts relevant to this situation:

1. PBL and the conventional medical curriculum lead to different ways of thinking about and dealing with medical problems.

A study conducted by Patel, et al. (1991) illustrates this phenomenon. These researchers compared 54 students (18 at each of three academic levels) at McMaster University in Ontario, Canada with 54 students (18 at each of three academic levels) at McGill University in Quebec, Canada. This study is fraught with many of the design weaknesses characteristic of the research in this field. Subjects come from systematically different, pre-existing groups (based on the entrance requirements of the two schools), and are not randomly selected or randomly assigned. Nonetheless, systematic differences emerge which seem highly related to the respective curricula.

The most notable difference, for my purposes, was that students from the PBL curriculum relied heavily on “backward reasoning” (in which they began with data, and developed explanations which adequately explain it). Because they gathered and assimilated information on their own, the PBL students tended to offer a variety of explanations for the clinical data.

In contrast, the students in the conventional curriculum (CC) tended to use “forward” reasoning, in which they compared the clinical data with previous clinical experiences, and essentially checked for a goodness of fit. These students tended to skip steps involving elaborate analysis of the data, and avoided providing alternative diagnostic explanations.

In general, PBL students generated many more explanations (and, therefore, many more erroneous ones) explaining the clinical data. CC students did not offer a variety of explanations, and therefore, offered fewer erroneous ones than the PBL students.

An important lesson to be learned from this, and similar studies comparing PBL and CC education (Norman & Schmidt, 1992), is that, by participating in different kinds of learning environments, students develop different ways of looking at lab data and generating hypotheses about what caused it.

2. Students like PBL.

There is evidence from a variety of studies in a number of settings that students find PBL enjoyable (Amos & White, 1998; Bligh & Wilkinson, 1997; Caplow, Donaldson, Kardash, & Hosokawa, 1997; Ferguson & Rutishauser, 1997; Itani, Miller, Church, & McCollum, 1997; Kaufman & Mann, 1996; Kindelan, Roberts-Harry, & Luther, 1997; Lancaster et al., 1997; Rand & Baglioni, 1997; Schwartz, Donnelly, Sloan, & Strodel, 1997; Van Treuren & Havener, ; Vasconez et al., 1993; Walton, Clark, & Glick, 1997; Wetherell, Mullins, Winning, & Townsend, 1996). In a study illustrative of this, Vasconez et. al (1993), and other researchers at the University of Kentucky (a medical school where PBL does not comprise the entire curriculum) collected data from 177 juniors at the completion of their junior clerkship. During this clerkship, students participate over a period of ten weeks, in activities ranging from PBL sessions (in which students met in small groups were given case-based problems, and went to the literature to resolve them) to traditional large-group lecture, to hands-on sessions in the operating room. This study evaluated three groups of students over three academic years, from 1990-1993. Students were asked to anonymously evaluate each of their learning activities on a Lichert scale ranging from 1 (poor) to 5 (excellent) over three categories: (a) quality of teaching, (b) importance of content, and (c) facilitated learning. In all cases, PBL ranked significantly higher than the other teaching/learning components in terms of student preference.

As shown above, the literature reflects the results of this study. In general, students enjoy participating in PBL.

3. Does PBL “Cover the Bases”?

One of the fears that many have of PBL is that, because PBL is relatively unstructured, students might not cover all the necessary basic science information taught in a conventionally structured curriculum (Bridgham et al., 1991; Kennedy & Wilkerson, 1993). In the context of the VMRCVM, this is not a particular concern. The College as a whole uses a fairly conventional curriculum, and, even in the Clinical Pathology course, which approximates elements of PBL, the concepts are

carefully predetermined and structured according to topic area, and basic concepts which accompany each case are taught.

However, this is an important issue that could potentially affect the design of the problem list generator. In short, the problem list generator is designed primarily to help students learn a process. But, if it can be shown that PBL, which makes no formal attempt to “cover” information, does, in fact address all the basic concepts in a conventional curriculum (and students learn that information), then it can be assumed that the clinical problem classifier would also be useful in “covering” information without an overt attempt to do so. Furthermore, since many popular conceptions of expert schemes of problem solving rely on a large knowledge base, it may be that the development of a knowledge base must be part of the problem list generator.

Opinion is divided on whether or not PBL covers all the desired information. One study found that students in a self-directed format identified 64% of the issues identified by the faculty before the class as important (plus additional issues considered important by the faculty that were not identified as such before the study) (Dolmans, Gijsselaers, Schmidt, & Van Der Meer, 1993). The authors were optimistic about this figure. And since PBL is the object of study on this issue, there is no indication whether or not students in a conventional curriculum would do any better. In general, there is evidence to suggest that students can be led towards essential basic knowledge through the problems presented. Kennedy and Wilkerson (1993) sought to determine whether or not different groups in a PBL curriculum, given similar problems, identify the same major issues related to the problems. They found that students in different groups all identified similar gross anatomical issues, even though faculty objectives (in terms of issues to be addressed) were not provided. The implication of this study is that students will identify and learn about basic general concepts without overt efforts to teach them, but through carefully designed “problems.” For our purposes, if the problem classifier uses carefully selected problems, it is likely that it will be able to impart essential concepts in the context of the problems being taught.

Knowledge Structures in Clinical Practice and Education

As has been shown in a previous section, there is certainly no shortage of models of human cognition. However, in this situation it seems particularly useful to examine general ideas about how successful clinical pathologists structure knowledge. Such an approach, however, must be tempered by Winn's warning, however, that "If we try to teach the skills of the expert directly to novices, we shall surely fail" (Winn & Snyder, 1996 p. 125). So, while knowing how experts think about clinical lab data should improve decisions concerning how to help novices approach problems like experts do, it seems more important to understand how the experts learned to approach lab data as they do.

Learner models of reasoning as they relate to curricula.

There is evidence to suggest that the way students acquire medical knowledge affects the way they organize that knowledge and use it to solve problems. For example, as seen previously, students using PBL are frequently observed engaging in hypothetico-deductive reasoning (or “backwards” reasoning) (Patel et al., 1991). Hypothetico-deductive reasoning, as defined by Merriam-Webster (1998), involves “making use of the method of proposing hypotheses and testing their acceptability or falsity by determining whether their logical consequences are consistent with observed data.” An examination of the assumptions and methods of PBL shed light on why a student who is essentially given a problem, a group of colleagues, and a library, would engage in hypothetico-deductive reasoning. The alternative would be guessing. Students in such a situation, lacking models, examples, and an appropriate mental schema, must attempt to find relevant information, hypothesize about why such information might affect their problem, and test the hypothesis(es). It seems reasonable to expect to find, as Patel, et. al. (1991) did, that in such a situation students would tend to generate a variety of hypotheses. To such students, lacking formal direction in terms of which hypothesis might be superior, any hypothesis is as good as any other unless proven otherwise.

Predictably, students in more conventional learning environments use more forward reasoning (Patel et al., 1991). That is, they compare current data with past experience, or existing frameworks of knowledge, and check for a goodness of fit. Such students have been presented with expert explanations of problems and their causes, and therefore have explanations for which to compare existing cases. These students, of course, can only diagnose problems they have been exposed to, and so, their ability to correctly diagnose in unknown domains is seriously hampered.

Interestingly, students who have experienced PBL, even after having developed experience and some frameworks of knowledge, tend to engage in hypothetico-deductive reasoning more than forward-thinking. Some researchers suggest that this is a potential weakness of PBL (or any other curriculum that encourages self-directed problem solving), for such a problem-solving model results in multiple hypotheses of diagnoses (including erroneous ones) and is a departure from what has been shown to be the medical “expert” model of reasoning (Patel et al., 1991). I will assert that this characteristic should actually be seen as a strength of the problem-based learning approach. For, as we will see, experts use backwards reasoning whenever they are not working in comfortable knowledge domains. Since novices have not developed mature and complete sets of knowledge for any given domain, for them to use backwards reasoning is an “expert” approach. Second, it is through the use of backwards reasoning that true experts avoid errors in clinical practice.

General models of diagnostic reasoning.

Research has revealed that different medical students and practitioners approach clinical problems in different ways depending on experience and expertise. There appear to be patterns in the way clinicians' thinking evolves over time, and there are a variety of ways to look at the way clinicians think.

Schmidt, et al. (1990) propose a stage theory of clinical reasoning which attempts to explain the available research. They suggest that clinicians employ four stages of reasoning which build upon, but do not replace, each other. Stage 1 involves the development of elaborate "causal networks." Stage 2 involves compiling the elaborate networks into abridged ones. In stage 3, the clinicians develop illness "scripts" and in stage 4, they develop and use "instance scripts." Basically, the progression from one stage to another occurs as practitioners gain practical experiences which provide cognitive representations of their knowledge. Causal networks of pathology are eventually replaced by a variety of instance scripts (past experiences with specific patients, symbolizing various medical problems). Clinicians will tend to use the highest stage of reasoning available to them, depending on the availability and complexity of knowledge they have in any given domain. A clinician who has experience with a previous patient that "fits" a current patient, will tend to use that experience as an instance script, whereas in an unknown situation, the physician will revert to earlier stages until finding one that best accounts for the current problem. While this theory explains the behavior of expert clinicians (and is consistent with the generally accepted concept of "automaticity" of expertise (Speelman, 1998; Winn & Snyder, 1996), its explanation of the underlying cognitive processes is lacking, and its implications for practice are disturbing. A method which compares previous cases with current ones, without verifying their accuracy, is fraught with possibility for error. As a matter of fact, practicing "expert" clinicians all too frequently misdiagnose for a variety of reasons which tend to involve careless consideration of data, including order of data presentation, ignoring relevant information, and inaccurate probability estimation (Bergus, Chapman, Gjerde, & Elstein, 1995; Christensen, Heckerling, Mackesy, Bernstein, & Elstein, 1991; Lyman & Balducci, 1994). For example, Christensen, Heckerling, Mackesy, Bernstein and Elstein (Christensen et al., 1991) report that medical practitioners will make diagnostic mistakes for no other reason than the framing of the problem (i.e. 95 % chance of survival is reported as opposed to 5 % chance of mortality.) Similarly, Bergus, Chapman, Gjerde, and Elstein (1995) found that family physicians erred in diagnosis as a result of whether or not the medical history was presented before or after the clinical data was presented. Therefore, while checking a current case against past cases is a good first step for the expert, the true expert (those who misdiagnose infrequently) must employ some mechanism for ensuring accuracy beyond a comparison, even in cases where the fit seems clear.

Bordage and Lemieux (1991) provide the framework for understanding how this process works. They compared the semantic structures of experts and novices as such structures relate to diagnostic thinking. Their study, based on structural semantic theory, provided clinical problems for a variety of experts and novices and examined their solutions in terms of semantic relationships. These semantic relationships were used to reveal the relationships identified by the participants that linked the various items of data in the clinical problem. Each relationship identified was referred to as a “semantic axis.” It was found that “the more diversified the diagnostician’s network of semantic axes, the better was his or her diagnostic accuracy.” In other words, the more valid abstract relationships the subjects found between bits of information in the problem, the more accurate their diagnoses. The number of semantic relationships for each case was not necessarily related to the number of words used to describe the analysis. In other words, an expert might appear to arrive at a diagnosis quickly and without much thought (with very few words), but a semantic analysis would reveal that the expert has understood and identified a great number of valid relationships between the data without spelling them out as such. Comparing this to Schmidt, et al.’s four stage model, it could be that part, perhaps most of the data about these relationships is embedded in what appear to the researcher to be mental “comparison” scripts, but the expert problem solver is aware of and does account for relationships between bits of data.

Bordage (1994) later examined his and Lemieux’s findings in terms of their implications for arriving at accurate diagnoses. In his words “The selection of an appropriate diagnosis is based on complex semantic connections (p. 885).” Such an assertion is more complete, and for the purposes of informing educational practice, more useful than Schmidt et al.’s (1990) assertion that “expert clinical reasoning is, to a large extent, based on the similarity between the presenting situation and some previous patient available from memory (p. 617).” The implication is that to truly become experts, students must become expert at associating relevant information within a case, rather than or in addition to comparing cases.

Several studies seem to support Bordage and Lemieux’s (1991) findings. For example Stevens (1991) employed a computer program which tracks students as they go through the problem solving process of forming a hypothesis, searching for relevant data, and arriving at a diagnosis. Students arriving at the correct diagnoses consistently chose hypotheses and tests which were consistent with each other and with the clinical problem. In the language of Bordage and Lemieux (1991), the successful students showed a higher number of correct semantic axes.

In a related study Cholowski and Chan (1992) examined diagnostic reasoning among nursing students. They used a variety of tests to measure each student in terms of a) content knowledge, b) logical reasoning (through expert evaluation of student performance on a diagnostic task using Biggs and Collis's SOLO taxonomy) c) nursing assessment (expert evaluation of each nurse's performance), d) nursing

diagnosis (whether the nurse arrived at an appropriate or inappropriate diagnosis) and e) approach to learning and information processing (using Biggs' Study Process Questionnaire). Several interesting results emerged. First, "higher quality reasoning" (in which students accounted for more intricate relationships between relevant items of information) showed a significant correlation to higher quality diagnoses. Path analyses further revealed that content knowledge contributed to logical reasoning which in turn contributed to a correct diagnosis. Information processing, in turn, had a positive effect on knowledge. Students whose information processing strategies were classified as "deep" (students focus on subject matter and relate it to previous knowledge) or "achieving" (strategies are aimed at success and involve time management and work space strategies) were more likely to have an understanding of the content than students whose study habits were classified as "surface" (strategies are aimed at reproduction and minimally meeting task requirements). In other words, students who really studied to understand the content, tended to understand the content, which in turn improved the measure of logical reasoning, which, finally, improved their diagnoses.

The implications of Cholowski and Chan's (1992) study fits Bordage's (1994) semantic axes concept nicely. It is clear that those who diagnose successfully have understood content and have organized it in terms of legitimate relationships between bits of data. The task, then, is to help students develop the knowledge structures of experts -- complete with content. There might be several ways to go about this. First, problem-based learning has been shown to work, though some research, as we have seen, suggests that this might not be seen an efficient model for students spend a great deal of time developing and testing incorrect hypotheses. In looking at this problem, Mandin, Jones, Woloschuk, and Harasym (1997) propose providing an expert scheme in the context of a problem when teaching diagnostics to students. They see this as a way to help students work with a problem with expert help, without ignoring content. One way to do this, they suggest, is to provide students with a possible diagnosis, and the data that lead to that diagnosis, and require the students to "fill in the lines." Another approach is to provide students with a logical scaffold (or "scheme") for organizing the data, and having them attach relevant data to that scheme. They argue (compellingly) that if students learn knowledge in the context of an expert scheme, or network, they will be more likely to remember the knowledge as it relates to other pieces of knowledge, and therefore more likely to solve problems by relating knowledge bits appropriately as proposed by Bordage and Lemieux (1991). Though their concept has not been tested in this context, it seems reasonable and consistent with general notions of learning as covered earlier. In the language of cognitive apprenticeship, their proposed expert scheme would provide a scaffold for learning.

Synthesizing: Theory and Research

A lot of ground has been covered, and a lot of concepts have been addressed. I think that, in a general sense, the pieces of the puzzle that have been presented can be

put together into a compelling case for certain prescriptions in dealing with a problem such as building the problem list generator. Those prescriptions follow:

Exposure to and practice dealing with medical problems should prove useful in the development of data classification skills.

- Experts remember and refer to previous problems when diagnosing -- learning in the context of a problem adds to this knowledge (Patel et al., 1991).
- Exposure to real life problems provides information and practice relevant to creating and building appropriate "semantic axes" to borrow from Bordage (1994). In the terms of cognitive psychology, we are schema building, in the terms of behaviorism, we are practicing behaviors in context, and in the terms of constructivism, our learning is situated.
- Students enjoy learning in the context of problem solving, and, therefore, are more likely to study the content using "deep" or "achieving" strategies, rather than "surface" strategies (Amos & White, 1998; Bligh & Wilkinson, 1997; Caplow et al., 1997; Cholowski & Chan, 1992; Ferguson & Rutishauser, 1997; Itani et al., 1997; Kaufman & Mann, 1996; Kindelan et al., 1997; Lancaster et al., 1997; Rand & Baglioni, 1997; Schwartz et al., 1997; Van Treuren & Havener, 1997; Vasconez et al., 1993; Wetherell et al., 1996).
- The use of authentic problems is one way to maintain an "active" learner -- which is prescribed by learning theory (Burton et al., 1996; Duffy & Cunningham, 1996; Winn & Snyder, 1996).

Any teaching of diagnostic problem solving cannot rely on a forward-thinking, or "goodness of fit" model for dealing with problems, but should rather rely on a model which forces students to deal with the process of synthesizing content .

- When experts are too quick to apply a previous case in solving a current one, or when they do not have a perfect match and try to force one, mistakes are the result (Bergus et al., 1995; Christensen et al., 1991; Lyman & Balducci, 1994).
- The domain for knowledge content for experts in the field of veterinary medicine (as is the case with many disciplines) is broad and ever-expanding. There is no way to ensure that any given case is equivalent to a previous one without engaging in analysis of data.
- Accurate diagnoses rely on an accurate network (semantic structure) of precisely related bits of information. Even experts have been shown to rely on such a structure, in a "concealed" way (in terms of semantic networks) when working within their domain of expertise and in an explicit fashion when working outside their domain of expertise. The metaphor is cognitivist, and works well with schema theory (Bordage, 1994; Bordage & Lemieux, 1991; Winn & Snyder, 1996).

- Even if a "goodness of fit" model worked for experts, such a model would have been built on an understanding of content (Bordage, 1994; Bordage & Lemieux, 1991; Winn & Snyder, 1996). (Through schema building, careful contingency/behavior building, or situated practice, depending on your theoretical persuasion (Brown et al., 1989; Skinner, 1954; Winn & Snyder, 1996)). To provide learners with "job-aid" type instruction without underlying principles, in a field where lives are in the balance and all possible problems encountered cannot possibly be explained without an understanding of underlying principles, seems foolhardy.
- Theoretically-based prescriptions for teaching cognitive skills require practice in context (i.e. Gagné's approach (1992), or a detailed teaching of all skills, sub-skills, and content (Merrill & Tennyson, 1977). Mandin et al. (1997) argue compellingly for a combination of the two approaches in which learners use an expert scheme for practicing cognitive skills.

- **A tool such as the problem list generator should not be designed or implemented without allowing for context.**

- Not only must the practice be contextual, as argued in the first point, but the instructional intervention must be based on a thorough "Tessmer/Richey-style" understanding of context (Tessmer & Richey, 1997).

- **A tool such as the problem list generator should be built on a thorough front-end analysis.**

- If a problem-based approach is used, problems should be designed carefully to address relevant content (Mandin et al., 1997).
- Regardless of the approach, good instruction will rely on a thorough understanding of expert performance and the process of gaining expertise in the given field (Burton et al., 1996; Duffy & Cunningham, 1996; Gagné et al., 1992; Merrill & Tennyson, 1977; Winn & Snyder, 1996).

- **A tool such as the problem list generator will allow time for learning and practice.**

- Regardless of theoretical orientation, the development of cognitive strategies, and high order intellectual skills, and the adoption and accommodation of new knowledge takes time (Burton & Magliaro, 1988; Lindsay & Norman, 1977).

- I would argue that, based on learning theory and the available research on the subject, a tool such as the problem list generator which is based on the above prescriptions is likely to prove effective. The next chapter will present how these prescriptions were used to design and implement the Problem List Generator.

CHAPTER 2: PRODUCT DEVELOPMENT

Chapter 1 outlined specific prescriptions to be used in the design and implementation of the proposed problem list generator (PLG) and provided a justification for those prescriptions based on relevant literature. This chapter will provide an overview of the design/development process, the resulting software tool, and additional relevant details in the process of designing, developing, and initially implementing and evaluating this tool.

Overview of the Development Process

The design and development of the PLG has taken approximately a year and a half, and late-stage debugging and enhancing are still taking place at the time of the writing of this document. The design process alone consumed the first 8 months of full-time work. Our first task was to paint an adequate picture of (or clearly define) expertise in this kind of clinical problem solving. This question led to the extensive review of literature reported in chapter 1. It also led to many hours of interviews with experts and learners in clinical pathology. Most of these interviews entailed giving a representative medical problem to an expert or learner, and watching and recording their process for dealing with it (while asking questions.) I emerged from these interviews with a pretty good idea of a process which I felt, if followed, would aid learners in developing the expertise we were seeking to foster. This is the process which is embodied in the PLG and will be described later.

Armed with an idea, supporting research, and ready access to expert knowledge, I began the design. Generally, I would work throughout the week and produce pieces of the instructional and/or interface design. I would report my progress in weekly meetings of the Medical Informatics Group. This group, which has changed somewhat during the passage of time, is made up of undergraduate students, veterinary students, graduate students, and professors with expertise in education, computer science, engineering, and clinical pathology. During these weekly meetings I would present my work, and receive feedback. These sessions frequently involved a great deal of give and take with diverse ideas, perspectives, expertise, and experience all melding into a continuously maturing instructional design (and later, instructional product). Difficult questions were frequently referred to the individual or group of individuals having the most expertise in the area most relevant to the problem. Where doubts remained, issues would be set aside until formative evaluation with real learners could shed additional light on the question.

With the instructional and interface designs nearing completion, we turned our attention to finding the right tool and person or group to implement the product. Research and expert advice led us to choose Java as the implementation tool. We then turned our attention to finding someone to do the programming. A number of groups and programmers looked at the project. The first group, collaborators from another department at the University pulled out of the project when it became clear that it was not as good a fit to their research agenda as they had thought. A second

programmer was hired, but became sick and had to withdraw from the project before completing any code. A third programmer designed and implemented the software, delivering it in a nearly debugged state. Personal circumstances forced him to leave the area, and a fourth programmer has been engaged to complete debugging and enhancement process.

Throughout the development process we conducted the formative evaluation (described in some detail later in the chapter). The weekly meetings continued, and evaluation plans, processes, and results continued to be reported and discussed in these meetings on a regular basis.

The development process was not without its exciting moments. The students to be used as participants in the evaluation were only available for a limited amount of time, and the evaluation was constantly being delayed by slowdowns and difficulties in all aspects of the implementation process -- from the Java code itself to the web server software, database, and intermediary server software. Also, as the semester wound to a close and schedules became more and more pressing, students became less and less eager to spend time assisting in evaluations. In the end, however, through cajoling, late nights, and many hours of hard work on the parts of all involved, the PLG was implemented and evaluated.

Overview of the PLG

The PLG is intended to accomplish the goals outlined in the beginning of Chapter 1. Namely, it is intended to aid students in the process of classifying and synthesizing clinical laboratory data, and to allow students to receive feedback for their problem lists and credit for completing them. This section is intended to describe how the PLG accomplishes these goals.

Appendix A of this document includes the full text of the Interface Design Document, which describes in detail how learners interact with the PLG. A full understanding of the interface is not necessary to understand how the PLG works, however. I will present here those screens and interactions which embody the instructional strategy of the PLG without the technical, precise, and laborious interface information available in the interface design.

The PLG is a web-based, case-based learning tool that does the following:

1. Allows students to access data including signalment/history/physical exam (SIG/HX/PE) information, and images for a variety of animals in the context of cases.
2. Allows students to record information relevant to the SIG/HX/PE that will later be used in their problem lists.
3. Presents students with relevant laboratory data for the animal in each case.
4. Requires students to identify and name data abnormalities.

5. Allows students to create a “problem list” using the information identified from the SIG/HX/PE and the data abnormalities that have been identified.
6. Requires students to submit a diagnosis.
7. Requires students to compare their problem lists and diagnoses to an expert problem list and diagnosis and evaluate their own problem lists based on the comparison.
8. Records in a database when students have completed problem lists and assigns credit for completing them if they have been completed on time.

The next several pages will provide a text-based description, with accompanying illustrations, of how the above eight steps are accomplished. Each set of clinical laboratory data that the students will deal with will be embodied in a case - a description of an animal with a certain set of symptoms and laboratory data. This use of cases is consistent with the course professor's current teaching method and is consistent with the idea from the review of literature that students should deal, as much as possible, with "real" problems.

Step One

Figure 2.1 is the Case Summary screen, which allows students to learn initial information about the animal to which the case pertains. As can be seen, an image of the patient is accompanied by a brief textual history. This screen is designed to provide the student with the essential information about the animal that would not be found in the laboratory data. Additional images and video can be added to accompany the case history if such images and video are considered instructionally useful for any given case.

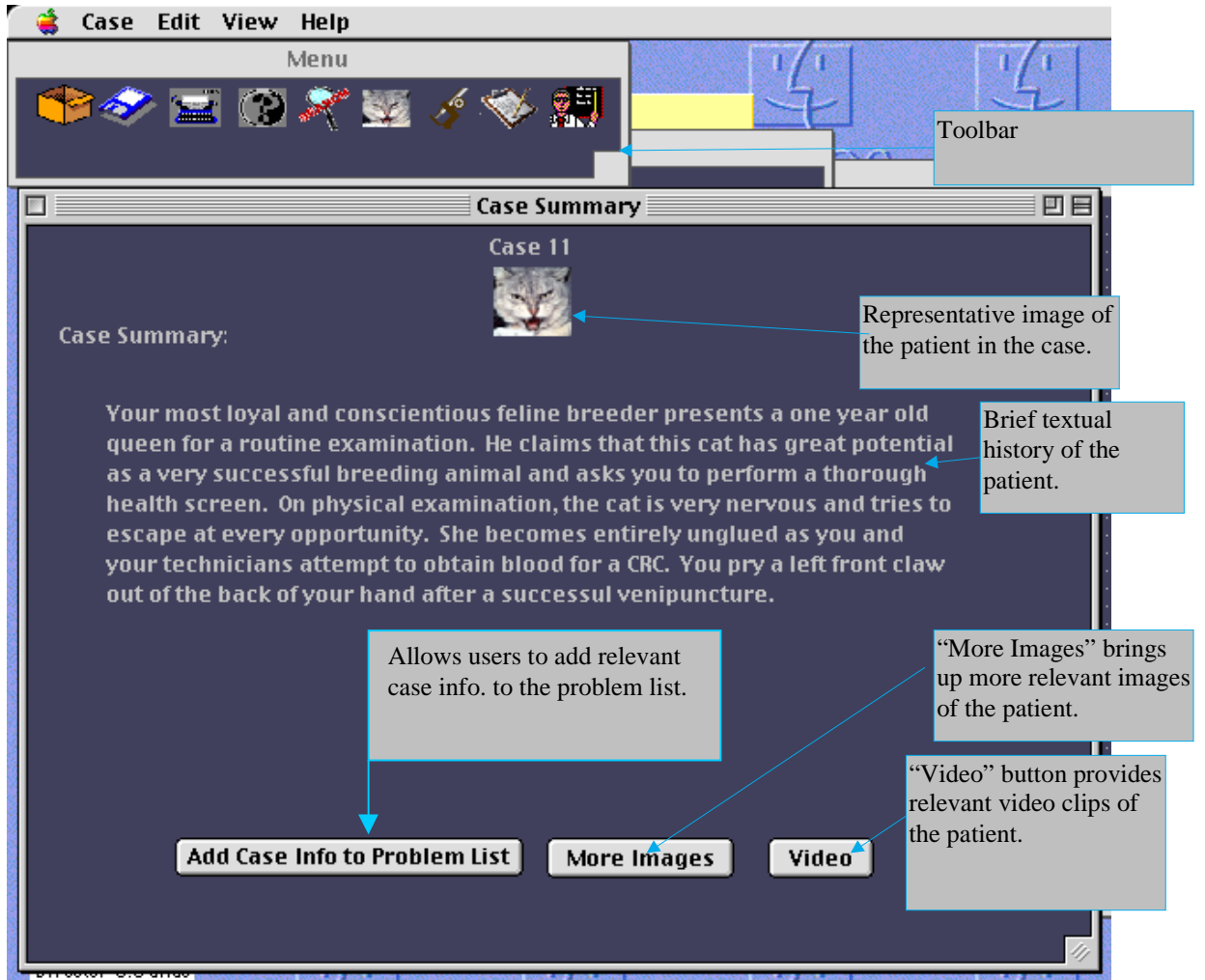


Figure 2.1. Case summary.

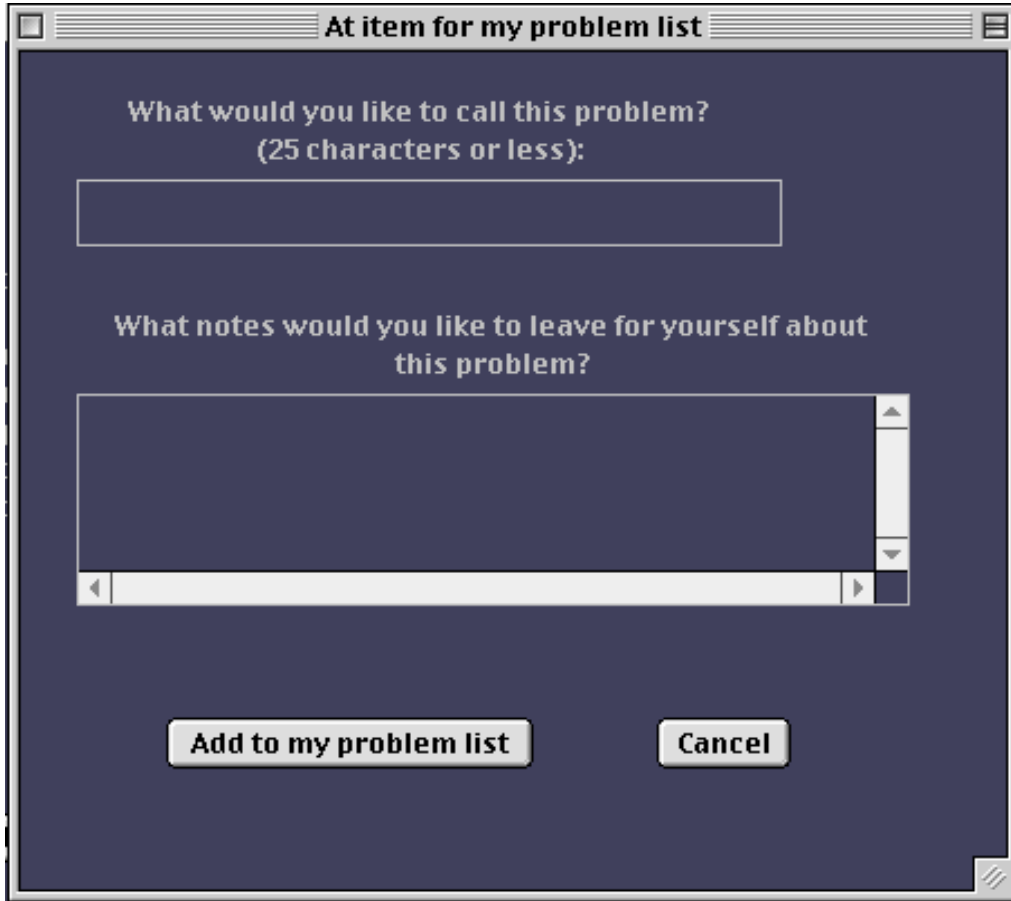
The image shows a dialog box with a dark blue background and a title bar that reads "At item for my problem list". The dialog contains two text input fields. The first field is preceded by the text "What would you like to call this problem? (25 characters or less):". The second field is preceded by the text "What notes would you like to leave for yourself about this problem?". Below the input fields are two buttons: "Add to my problem list" and "Cancel". The dialog box has a standard Windows-style border with a close button in the top-left corner and a maximize button in the top-right corner.

Figure 2.2. Adding an Item to the Problem list from the lab data screen button

Step Two

When learners identify information in the Case Summary that they feel will prove useful to them in generating a problem list, but will not be likely to appear in the lab data, they will be able to enter that information by using the "Add Case Info to Problem List" button. Doing so causes the screen shown in Figure 2.2 to appear. Once the user has entered a name for the problem or phenomenon identified, and has written a description of it (if desired), clicking the "Add to my problem List" results in the problem appearing in the "Data Abnormalities I've found" column of My Problem List" (as shown in Figure 2.5).

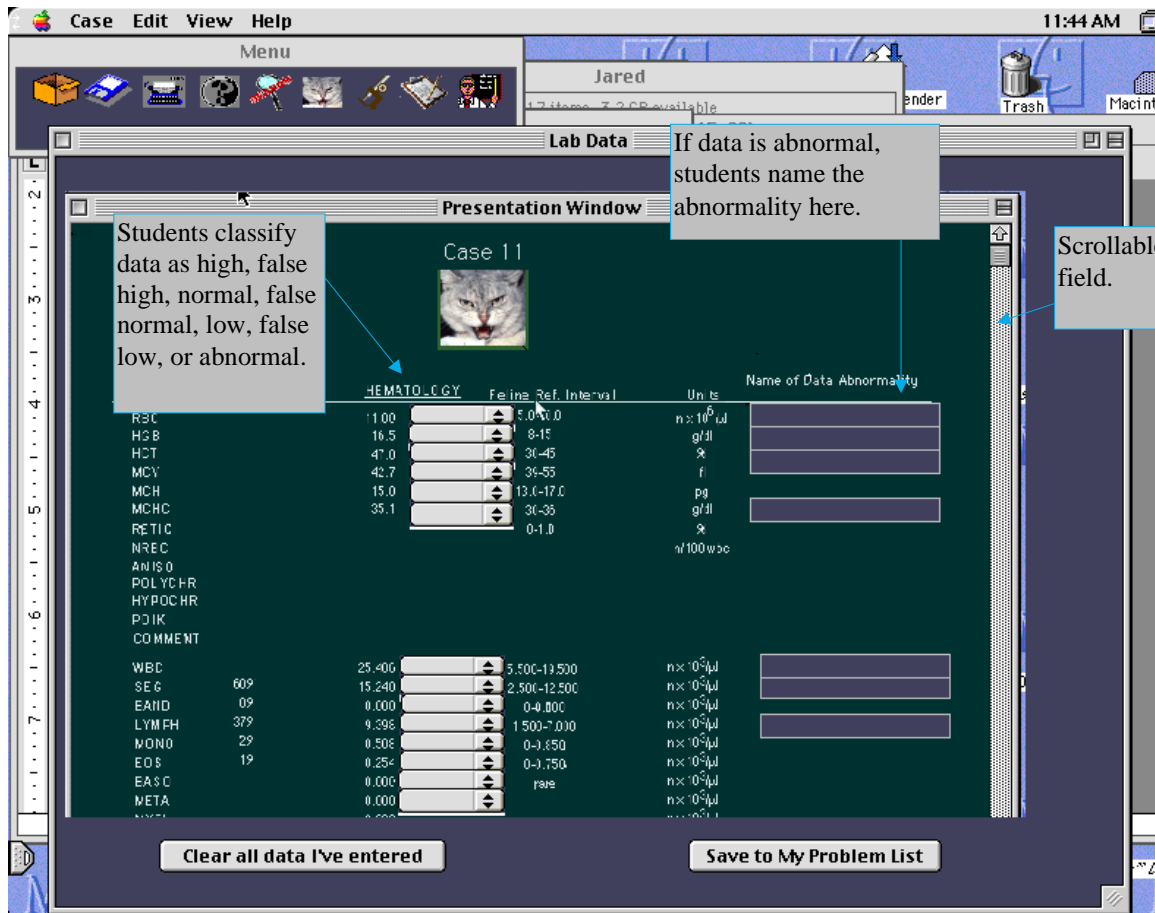


Figure 2.3. Lab data.

Steps Three and Four

The next interaction of instructional consequence that will be available to the learners is illustrated in the data classification screen. Here the learners initially encounter the laboratory data. This screen can be seen in Figure 2.3.

The data classification screen is designed to allow students to classify each item of laboratory data (normal, high, low, etc.) This is accomplished by manipulating the small pull-down menus appearing to the right of each data point. If data points are changed to anything other than normal, a text box such as the ones shown in the right column of Figure 2.3 will appear. In each text box the learner will enter the name of the corresponding data abnormality.

If a data abnormality name is entered which does not match the appropriate name in the database, a dialogue box appears indicating that the name entered is not correct. At the third unsuccessful attempt to name the data abnormality, the same dialogue box will appear, only this time, it also includes a button labeled "Tell me." If the "Tell me" button is clicked, the dialogue box will disappear, and another will

appear stating “When [name of data] is [high or low] it is known as [name of data abnormality].” Checking the students at this point is important for several reasons. First, if the students proceed to the problem list with incorrectly identified data abnormalities, such a mistake would probably lead to problems in generating an accurate problem list. Also, since knowing the names of data abnormalities is an instructional objective, and since it is feasible for the system to check this knowledge, it seems logical to do so at this point. Students are given the correct name of the data abnormality after three unsuccessful attempts, on the one hand to reduce the likelihood that they will simply allow the system to do the work of naming the data abnormalities for them, and on the other hand to make sure that they are not prevented from completing the exercise for lack of remembering how a name is spelled.

The lab data screen is purposefully laid out to reflect the printouts that veterinary practitioners are forced to work from in the field. Learning theory suggests that the more a practice environment approximates the environment in which the task will have to be performed, the more likely the knowledge and skills gained are to transfer to the work environment.

The system will not allow learners to begin a problem list until all identified data abnormalities are named correctly. This will ensure that the learner does not begin a problem list without all the necessary pieces for completing it (as identified by that learner).

Step Five

Once the learner has identified and named all the relevant data abnormalities, the problem list screen, as shown in Figure 2.4 will become available. Any time a data abnormality is named (Figure 2.3), or a problem is added to the problem list from the screen shown in Figure 2.2, the data abnormality or problem identified appears in the column on the right hand side the "my problem list" screen (Figure 2.4). Students create a problem list by dragging data abnormalities from the column on the right (Figure 2.4) to the box on the left and manipulating them there. In the box to the left (Figure 2.4) students can enter mechanisms, re-label abnormalities as mechanisms (if they function as such), add notes in text boxes attached to each mechanism or abnormality, and arrange mechanisms and abnormalities in relation to each other. The problem list communicates causal relationships between items of data which are arranged in an outline format. That which is above and to the left causes that which is below and to the right. Abnormalities on the same level are both caused by the same thing, but do not affect each other.

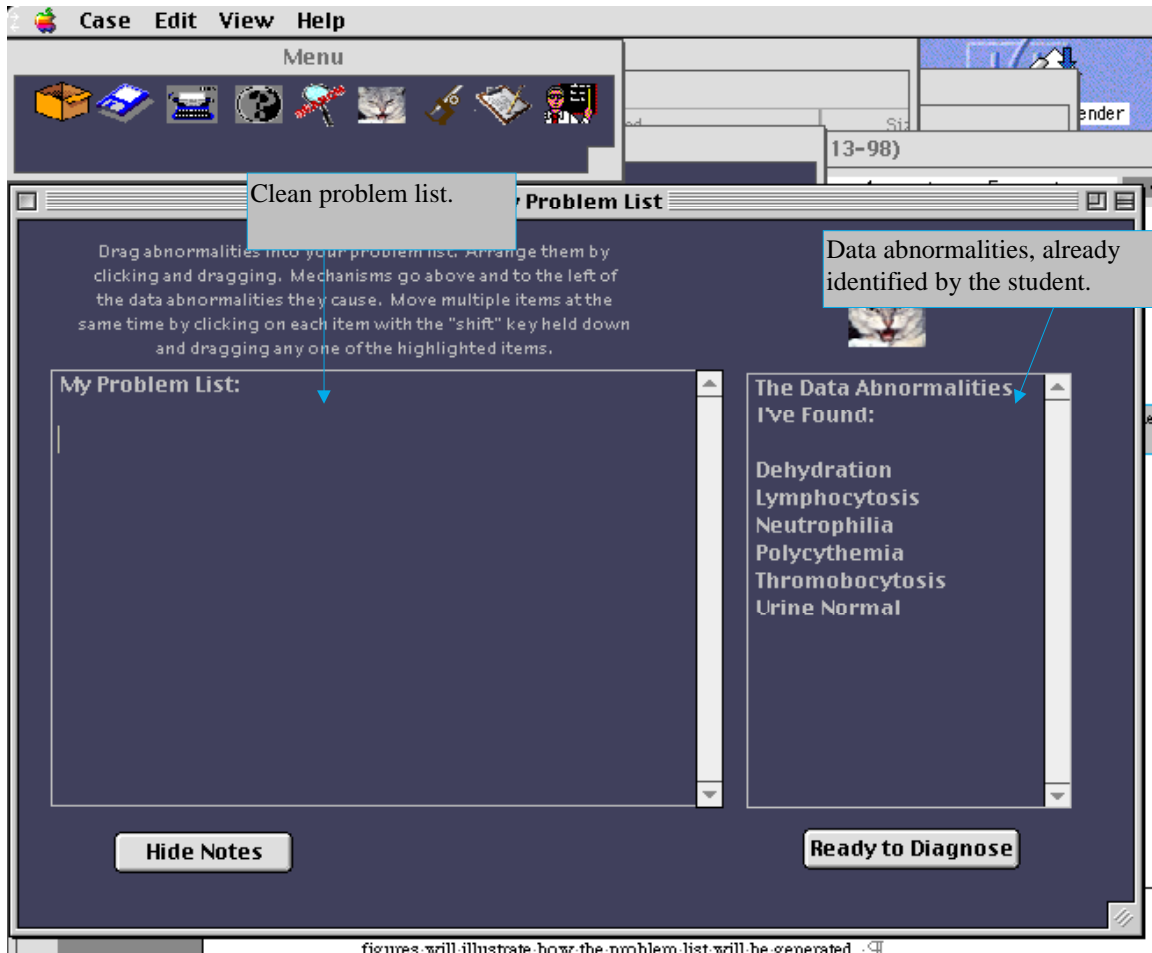


Figure 2.4. The learner's problem list.

Figure 2.5 illustrates a completed problem list. This problem list communicates (among other things) that the mechanism "Splenic Contraction" causes polycythemia and thrombocytosis. The notes fields for "difficult phlebotomy" and "cat is struggling" are visible while the other notes fields are not. This screen is designed as a "scaffold" of sorts for the learners. It allows them to see at once all the data abnormalities they have identified and requires them to organize them all in a meaningful way (one which identifies causal relationships.) Thus, this screen, while not providing the answers, is designed to force the learners, in a way, to ask all the right questions.

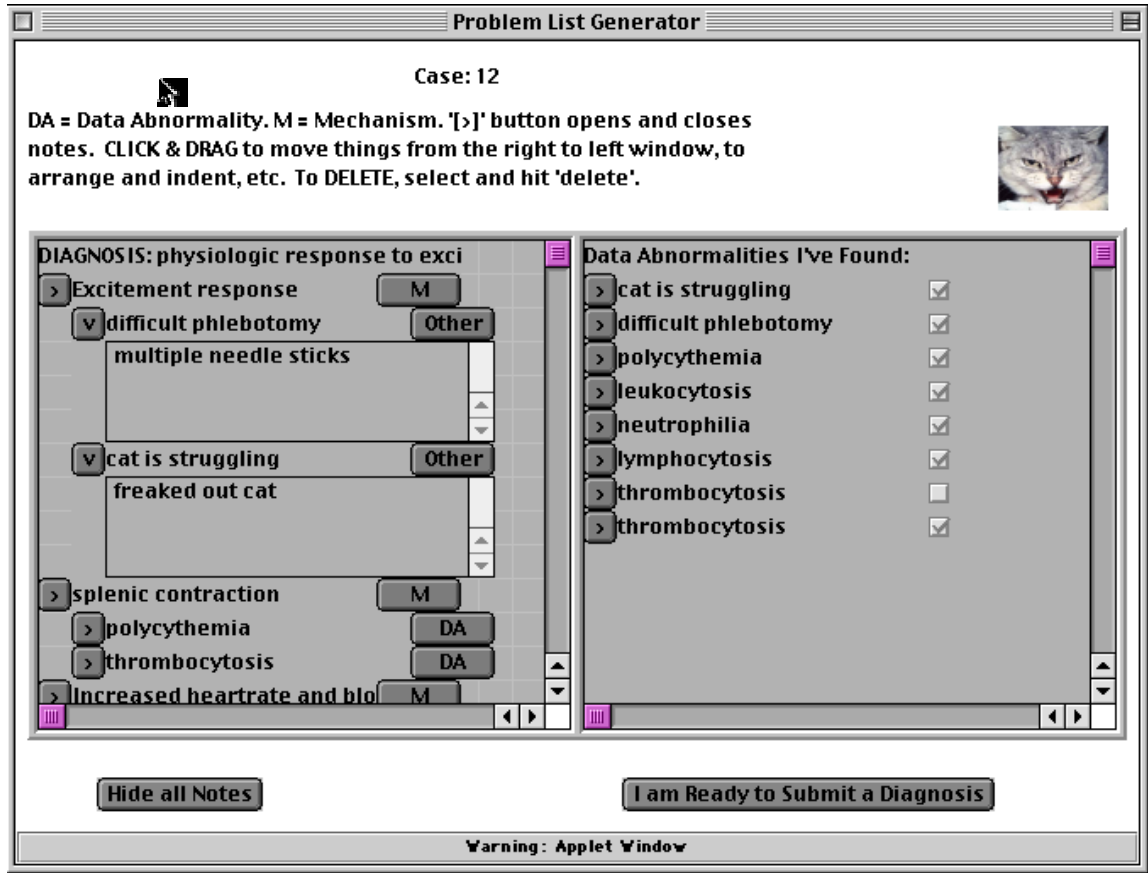


Figure 2.5. Completed problem list.

The three screens explored up to this point are all available to the learner throughout the process of generating the problem list. This is an iterative process. Students might decide that what had appeared to be "high" data, is actually "false high", etc., and will be able to go back and change it in all relevant screens accordingly.

Steps Six and Seven

When students have created a problem list to their satisfaction, they can submit a diagnosis and compare their problem list to the expert problem list. To prevent the students from simply copying the expert problem list instead of going through the process of generating their own, their problem list will no longer be able to be modifiable once they have submitted their diagnosis (which makes the expert problem list available).

The expert problem list can be seen in Figure 2.6, below. As shown, this screen also displays the learner's problem list so that the two may be compared.

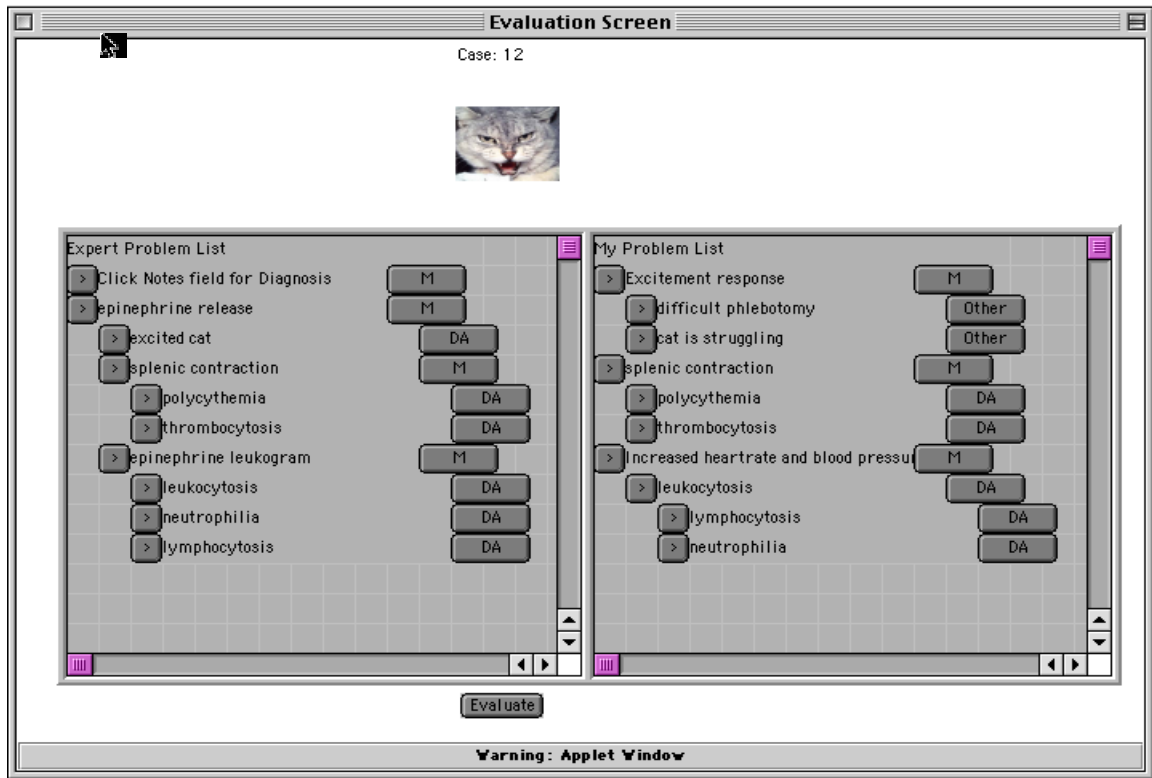


Figure 2.6. “My Problem List” and “Expert Problem List” compared.

This screen is designed to answer the questions that learners might not have been able to answer while generating their own problem lists, and to reinforce for learners that which they had identified correctly. Here the learners can read the expert solution, complete with notes, and compare their solution with the expert solution. This screen is designed to provide the learners with meaningful feedback on their efforts, fill in knowledge gaps that may have been missing, and prepare the learner for the lecture in which the professor will answer questions on this problem. (Learners will be able to print their problem lists or the expert problem lists for study or review purposes.)

As seen in Figure 2.7, once they have seen the expert problem list, learners will evaluate their own problem list on a scale of one to five for the purpose of reporting how closely it matches the expert problem list. This self-evaluation will not be used for grading purposes, but will be available to the professor for an indication of how each student feels he/she is doing.

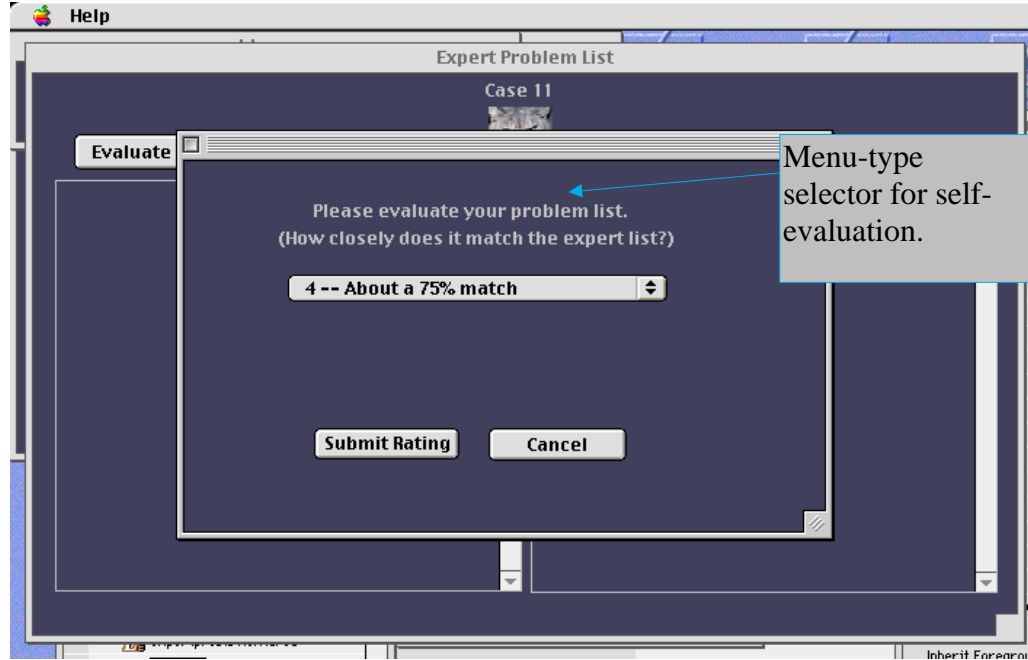


Figure 2.7. Evaluation dialogue.

Step Eight

Once the students have completed this process, their completed problem lists and self evaluations are recorded in a database, and they are given credit if the case has been completed on time.

Summary of the PLG

The PLG provides what might be seen by Mandin , Jones, Woloschuk, and Harasym. (1997) as an "expert scheme" and problems to which the scheme might be applied. In this case, the problems are cases which increase in length and complexity as the semester progresses. The expert scheme requires students to deal with the data in sizeable chunks, first by viewing a case history, then by classifying data and identifying and naming data abnormalities, and finally by arranging all the aberrant data in a causal structure. The strategy also allows students to compare their solutions with an expert solution.

Designing the PLG

The complete interface design of the problem list generator is available in Appendix A and the complete instructional design in Appendix B. This section is intended to merely report the process and tools used in the design, as well as the reasoning that led to the design and final product.

Interface Design

Since the PLG is a computer-based tool, it seemed clear that the design of the interface would be an important aspect of the instruction. The “nuts and bolts” of the interface design were based on Hix and Hartson’s “Designing User Interfaces” (Hix & Hartson, 1997). Therefore, “look and feel” decisions having to do with button placement, screen size, colors, etc., were made in an effort to make basic navigation, etc., as intuitive as possible. In terms of getting around in the program, our hope was that the software would be relatively transparent. My original thought was to do completely separate instructional and interface designs. In doing my instructional design, however, when I reached the point of articulating the strategy, it became very clear that the interface was the manifestation of the instructional strategy. Therefore, the instructional and interface designs were kept separate up to the point of articulating the strategy, and then merged.

A pull-down menu was selected as the program’s basic navigational tool. Accompanying the menu is a toolbar containing some of the program’s most common features. These features were intended to be as “transparent” as possible, and therefore, were based, as much as possible, on convention.

Many other “interface” design decisions were actually instructional in nature, and at times the interface and instructional goals appeared, artificially, to be at odds. The design of the PLG, for example, required students to enter, from memory and spelled correctly, data abnormalities with names like “thrombocytosis” and “hyperaminotransferasemia”. One interface design expert on the team immediately called the rest of the team to task, contending that requiring text entry is a bad idea in general, and represents a clear violation of human memory limitations guidelines. She suggested that we could make our interface much more usable if we provided students with a pull down menu of possible choices, or, better yet, had the system simply provide the names of the data abnormalities for the students. While apparently correct from a usability standpoint, the suggested change would have undermined the instructional objectives, which involved (among many other things) the students’ ability to produce these data abnormality names from memory. When this was pointed out to the interface design expert, she agreed that, in this instance, the best interface is one that appears to contradict accepted general norms of interface design. While it requires cognitive effort of the users, that cognitive effort is planned and productive. The power of combining interface and instructional design is that it allows us to ensure that the effort involved in using the software we design is planned and productive, and occurs, as much as possible, when and where we intend it to.

Instructional Design Models Employed

I used Dick and Carey’s “The Systematic Design of Instruction” (1996) for the initial instructional design. I find this model to provide a good basic framework

for approaching instructional problems, and used it as such. Because the review of literature called for extra attention to be given to the contextual analysis (and specifically, the front end analysis), I bolstered that process with a model for performing contextual analyses taken from Tessmer and Richey. This process explores three different “contexts” at three different levels (1993). I again turned to Tessmer (1993) for guidance in performing the evaluation. (Though the process prescribed by Tessmer is nearly identical in form to Dick and Carey’s process, it is more detailed.)

The Development Process

After the initial instructional and interface designs were completed, a paper prototype of the problem list generator was created. This was used to test the initial usability of the design. After the paper prototype was created, used, evaluated, and refined, a programmer was hired to implement the problem list generator. Because of the demands of the program, it was determined that it would be implemented in Java. It took the programmer approximately six months to complete a version of the software that could be evaluated, and, as of the time of this writing, the program is still being debugged.

Initial Evaluation Efforts and Modifications

The Expert Review Process

The Problem List Generator was designed and built with the input of a multidisciplinary group (The Medical Informatics Group) comprised of professors and graduate students in Clinical Pathology, Medical Informatics, Instructional Technology, and Computer Science. This group met weekly to discuss the group’s various projects, including the Problem List Generator. The instructional design (up to the point of specifying strategy) for the Problem List Generator was completed in early 1998 (between January and March). The first version of the combined instructional/interface designs was presented to the full group in late March of 1998. A design judged ready for implementation was finalized in September of 1998 (though that design was further altered after initial interface testing). Between March and September of 1998 the interface design underwent 26 revisions. All of the revisions that occurred before late August were based on expert feedback from the experts in the Group. Revisions throughout September were based on expert opinion and data gleaned from the evaluation of the paper prototype (which will be discussed next.)

The paper prototype

A paper prototype is a set of screens and menus, built on paper, and designed to support certain common tasks that represent the functionality of the program as a

whole (“benchmark tasks”) (Hix & Hartson, 1993; Hix & Hartson, 1997). A person designated as the “computer” runs the prototype as the participant performs the benchmark tasks. In this case, the participant was given a pen to use as both cursor and keyboard. The participant would “click” or “double-click” on buttons and menu items, etc. with the pen, as well as using it to enter text into text fields.

Evaluation of the paper prototype

Participants

The paper prototype was evaluated with three learners, one considered advanced, one considered average, and one considered below average. In this context, advanced, average and below average refer to the student's ability to classify and interpret lab data, and not to the student's class ranking or performance in previous classes. Students were evaluated and categorized by Dr. Holly Bender in an interview setting where she and each student discussed a specific case. These interviews are a regular part of the course. The categorizations were kept confidential. Students participating in the one-to-one evaluation of the paper prototype sessions were selected from a pool of volunteers who had been evaluated by Dr. Bender.

Observation

Students participating in the one-to-one evaluation of the paper prototype were observed and videotaped. Because the paper prototype was not “fully functional” it was used primarily to measure usability. Translated into Dick and Carey's (1996) language, the paper prototype was generally used for identifying problems with “clarity.” In other words, in this evaluation, we were primarily looking for whether or not the students could figure out how to “get around” in the PLG, and whether or not they liked the look and feel.

For the evaluation of the paper prototype, I borrowed and adopted a procedure designed by Hix and Hartson (1993; 1997) for testing interfaces. This procedure involves identifying and testing “usability specifications.” In this process, I identified key “benchmark” tasks designed to test every function of the system. Each participant was observed performing each benchmark task and one or more attributes such as time taken to complete a certain task and number of mistakes for first time use were recorded. Descriptive data was also recorded, particularly when learners encountered “critical incidents” -- moments at which it was difficult or impossible for the learners to know what to do next. Quantitative results were recorded on a “Usability Specifications” table as seen in Table 2.1. Other observations were recorded on a separate data collection sheet (see Table 2.2) and mapped to the appropriate tasks. In this particular context, Hix and Hartson's (1993) method was altered slightly. Time on-task was recorded, but time “manipulating” the interface was separated from time spent dealing with the instructional tasks of classifying lab data or generating a

problem list. (This was more of an issue later on with the completed program.) Also critical "learning" incidents (moments in which the learners are stumped by the learning problem) were separated from interface-related critical incidents.

In Table 2.1, the first column, "Usability Attribute" refers to what is to be measured. In most cases the "initial performance" (i.e., the first time the user performed the task) was reported. However, as can be seen, other attributes, such as "initial satisfaction" and "learnability" were also measured. The measuring instrument column refers to the specific benchmark task given to the learners. Column three represents the value to be measured -- most commonly either time taken to complete the task, or number of mistakes that occurred while completing the task. The fourth column indicates the "current level" (generally the level at which users can use a similar program already available. Because there were no similar programs available for comparison, the current level was the amount of time it took an expert to complete each task using the paper prototype. The fifth column "target level" represented the performance goal. In the case of timed goals, target level was arrived at by measuring how long it took an experienced user to complete the task and multiplying that number by 2. The other target levels were arrived at a little more arbitrarily, but were, in all cases, intended to be rigorous enough to catch major usability problems. Column six represents the results obtained from the participants

Learners were encouraged to talk as they worked with the PLG, and the observer asked questions when it seemed necessary, to understand what the participants were thinking or feeling. At the conclusion of the paper prototype evaluation session, the observer asked each participant to complete a brief questionnaire including lichert-type indicators of their experience using the system. This questionnaire was selected after the expert review process, is based on Shneiderman's (1992) usability questionnaire and was designed to provide an overall indication of the program's usability.

Table 2.1

Usability Specifications: Evaluation Results -- 3 User Mean -- Paper Prototype

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	OBSERVED RESULTS (mean)
#1 Initial Performance	“Get into the system” per benchmark 1	Length of time taken to successfully complete task	11 SEC	22 SEC	8 SEC
#2 Initial Performance	“Open Case 11” per benchmark 2	Length of time taken to successfully open Case 11	40 SEC	80 SEC	240 seconds for 1 user. Average 28.5 sec. for other 2
#3 Initial Performance	“Print My problem List” per benchmark 3	Length of time taken to successfully complete the task	73 SEC	146 SEC	37 seconds
#4 Initial Performance	“Find out” per benchmark 4	Length of time taken to successfully complete the task	NOT AVAIL	30 SEC	Show stopper for 1 user Ave 17.5 other 2
#5 Initial Performance	“Find, mark and label data” per benchmark 5	Length of Time taken to successfully complete the task	100 SEC	3 MIN, 20 SEC	Wide variation due to the nature of the task -- Max 5 min, Min 2:14
#6 Initial Performance	“Create a problem list” per benchmark 6	Length of Time taken to successfully complete the task	NOT AVAIL	5 MIN	137 Min (Also fairly wide variation due to the nature of the task)

Table 2.1. (continued)

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	OBSERVED RESULTS (mean)
#7 Initial Performance	“Create a problem list” per benchmark 6	Number of errors observed while completing the task	0	<2	0
#8 Learnability (after 10 minutes of practice)	“Create a problem list” per benchmark 6	Length of time taken to successfully complete the task	NOT TESTED	NOT TESTED	NOT TESTED
#9 Learnability (after 10 minutes of practice)	“Create a problem list” per benchmark 6	Number of errors observed while completing the task	NOT TESTED	NOT TESTED	NOT TESTED
#10 Initial Performance	“Create a note” per benchmark 7	Number of errors observed while completing the task	0	<2	0
#11 Initial Performance	“Create a note” per benchmark 7	Time taken to complete the task	27 SEC	1 MIN	2 users done with bench-mark 6 - other: 41 sec
#12 Initial Performance	“Submit diagnosis” per benchmark 8	Time taken to complete the task	NO DATA	1 MIN	40 SEC
#13 Initial Performance	“Report” per benchmark 9	Number of errors observed while completing the task	0	<2	0
#14 Initial Performance	“Report” per benchmark 9	Time taken to complete the task	20 SEC	40 SEC	46.6 SEC

Table 2.1. (continued)

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	OBSERVED RESULTS (mean)
#15 Initial Performance	“See if word appears” per benchmark task 10	Number of Errors observed while completing the task	0	<2	0 for 1 user 5 for 1 user show stopper for 1 user
#16 Initial Performance	“See if word appears” per benchmark task 10	Time taken to complete the task	17 SEC	34 SEC	17 sec for 1 user 55 sec for 1 user show stopper for 1 user
#17 Initial Performance	“Close Case 11” per benchmark task 11	Time taken to complete the task.	13 SEC	26 SEC	17 SEC
#18 Initial Satisfaction	Questionnaire	Average scores from all items on the Questionnaire. (Includes items #4, 6, 7 10, 11 and 13 of Quis Questionnaire)	NO DATA	8	8.3
#20 First impression	User reaction	Number of negative/positive remarks during session	NO DATA	POS > 5 NEG < 5	Overall favorable (9 positive; 3 negative)

Table 2.2

Data Collection Sheet

TASK NAME: PLG	Participant ID: 01 Task Start Time: Time to Perform Task: No. of errors:		Date: Task End Time:
Critical Incident Description	Tape Counter	Evaluators Comments	
1. #2 The screen didn't open #3 Terminology		#2 Technical error -- fix. #3 History vs. Sig/Hx/etc.	
2. #5 Screen was too small/lost data on closing #6 Closed program accidentally		#5 -- Technical error -- we need to fix that one (in terms of losing data) Do not know about the screen being too small. #6 Here she meant to get I.E. out of the way, and closed it (thereby closing the PLG as well.)	

Note. (Taken from (Hix & Hartson, 1997))

Based on the results obtained from the evaluation of the paper prototype, certain changes were made to the interface design. The revised interface design was given to a group of programmers who, as part of a strategic alliance that was in negotiation between them and the Medical Informatics group, were going to implement the Problem List Generator. When this group realized that they would be expected to implement the interface as it had been designed, however, they said that it would be too difficult, and withdrew from the project. A second programmer was located and engaged to do the project, but became sick after several months and before being able to make any substantial headway. In November of 1998, a third programmer was located who agreed to implement the PLG as designed, and who completed a version of the PLG whose core components could be evaluated by March of 1999.

One-to-one Evaluations of the PLG

In March of 1999, the “core” of the PLG was evaluated in a series of one-to-one evaluations. This “core” implemented the PLG as designed (and included the changes that emerged from the evaluation of the paper prototype), with some limitations. The most notable of these limitations were that a) the database interaction was not working correctly and therefore it was only possible to do small cases with the PLG b) “secondary” features, such as the “print”, “help”, and “search” functions had not yet been implemented, and c) the same bug that prevented large cases from working, also prevented the students’ completed problem lists from being saved correctly. It was decided to proceed with the small group evaluations in spite of these limitations because most of the benchmark tasks were supported by the PLG at this point, (at least in the interface), and because time was running out in the semester.

Participants

The one-to-one evaluation was carried out with six learners, one considered advanced, one considered average, and four considered below average. These students were enrolled in the Clinical Pathology course at the time of the evaluation. As was the case with the evaluation of the paper prototype, advanced, average and below average referred to the student's ability to classify and interpret lab data as determined in an interview setting by Dr. Holly Bender. Participants in the one-to-one sessions included all the students from a pool of volunteers who had been evaluated by Dr. Bender at the time the one-to-one evaluations were conducted. Fortunate happenstance led to the proportion of advanced, average, and below average learners who participated. It was the opinion of the committee that below average learners would be the most likely to be affected by changes to the instruction, and therefore, that if any group were to be “over represented” in the one-to-one evaluations, it should be those who were considered below average.

Observation

Students participating in the one-to-ones were observed and videotaped. Dick and Carey's (1996) three categories, "clarity", "impact" and "feasibility" were used as general categories for observation during these one-to-one sessions.

Clarity.

The same method of assessing clarity that was used for the paper prototype was used for the one-to-one evaluations of the PLG. (Benchmark tasks were given to each learner, and the results were recorded on a Usability Specifications table.) Other observations were recorded on a separate data collection sheet (example in Table 2.2) and mapped to the appropriate tasks.

Feasibility.

Results of the procedures for testing clarity also proved useful for the initial evaluation of feasibility. Tasks that learners were completely unable to perform, or that took an excessive amount of time were noted and the causes were identified. At this point, some bugs in the software that had not been discovered were discovered. It also became clear at this point that some computer users were not sufficiently well acquainted with the operating system being used (Windows 95) to use some of the navigation features that had been built into the PLG. This evaluation was conducted in a quiet, controlled environment, so it was difficult at this point to determine the feasibility of the using the PLG in other environments such as home or a busy computer lab.

Impact.

It was hoped that two indicators of impact would be available following the one-to-one evaluations. The principle of these was to be the problem list generated by the learner. The accuracy/completeness of these problem lists was to be evaluated by a subject matter expert. Unfortunately, as mentioned earlier, a bug in the PLG prevented complicated cases from being used with the PLG at this point. Therefore, students used the PLG with small, easy cases, and cases to which they had already been exposed. This prevented an accurate measure of impact for these evaluation sessions.

Chapter 2 has reviewed the design, development, and initial evaluation of the PLG. Chapter 3 describes the methods and results for the small group evaluation and field trial of the PLG.

CHAPTER 3: EVALUATION METHODS AND RESULTS

This chapter will report methods and results for the small group evaluations, and for the field trial. As was the case with the expert review and one-to-one evaluations which were described in chapter 2, the primary purpose of these two evaluations was to determine the clarity, feasibility, and impact of the PLG.

Small Group Evaluation - Methods

Participants

A total of 20 people participated in the small group evaluation sessions. All participants were enrolled in Clinical Pathology at the time of the small group evaluation. Of those who participated, 14 were female and 6 were male. Participants ranged in age from approximately 20 to approximately 40 years of age with the bulk of the participants being in their early 20's. In terms of age and gender, the group was very representative of the students at the Virginia Maryland Regional College of Veterinary Medicine. (Of the student body, 70% are female, 30% are male, and most are in their early 20's.) The participants were not randomly selected, but were unpaid volunteers. Some effort was made to recruit females to participate, as a greater group of males (proportionally) than females originally volunteered to participate. Because all available volunteers were used for the small group evaluations, it was not possible to select by ability level.

Procedures

Four small groups participated in the small group evaluation. The small group evaluation sessions occurred in a small computer lab housing five computers running Windows 95 and several other computers running the Macintosh OS. Only the Windows 95 machines were used for the small group evaluation, and, during the evaluation sessions other lab users entered and exited the lab occasionally to use the Macintosh computers. The noise produced by other lab users was minimal.

One hour was allocated for the small group sessions, but most took about 1 hour and 15 minutes. At the beginning of each small group session, the participants were given an informed consent form to sign (Appendix C), and a directions sheet (Appendix D). They were asked to read and sign the informed consent form, and to read the directions sheet as well. Some participants chose to read the directions sheet, and some appeared to only give it a perfunctory glance. All read and signed the informed consent form. Directions for the students were placed on the white board at the front of the lab. The directions were as follows:

1. Login. Name = your P.I.D. Password = your grade password
2. Open Case [the case number].
3. Look at the Sig/Hx/PE and make appropriate notes.
4. Classify and name the laboratory data.
5. Make your problem list.
6. Submit your diagnosis.

7. Compare your problem list to the expert list and evaluate it.
8. Fill out your questionnaire.

The observer briefly reviewed these directions with the students, and then asked them if there were any questions. After answering questions, the observer told the students they could begin. The students were asked to complete a case which was a current homework assignment for them, and were told that they would receive credit for completing the case using the problem list generator. The students were also told that they could ask questions at any point, and that they should feel free to interact with each other if they wished. The participants then worked through the assigned case, some asking questions occasionally.

About 10 minutes before the allotted hour was up, the students were asked to end their problem lists (whether or not they were complete) and move on to the next step (submitting a diagnosis). This was done in an effort to keep the sessions to the promised one-hour length. After all the students had finished their sessions with the PLG, they were asked to complete the questionnaire. When all students had completed the questionnaire, an audio taped focus group session was held. The focus group sessions were informal discussions in which the observer asked the group of students for their general reactions to the PLG, and whether or not they saw it as a viable alternate to the way they currently did their problem lists (on paper).

In all cases but one, each student used an individual computer. In one of the four sessions, however, due to a communication error, six participants came (only 5 computers were available.) Rather than send a volunteer away, the observer allowed two participants to work together on one computer. Since students were already allowed to collaborate, this was not seen as being a major threat to the validity of the data gleaned from these participants. During the small group sessions, the observer took notes when learners encountered particular difficulty, and answered questions as needed.

During the small group evaluations a new bug was discovered. While the PLG could now be used to work large and complex cases, the completed problem lists were not being saved to the database correctly. This meant that a large and complicated problem list might be saved with only one or two lines of text intact. This did not affect the use of the program at all from the users' perspective, but it did affect the ability of the developers to assess impact, as will be shown hereafter. There were also several known bugs that had not been fixed at the time of the small group evaluations. These bugs were pointed out to the participants, and work-arounds were provided.

Data Sources

Introductory Note

There were four small group evaluation sessions, one with four participants, two with five participants, and one with six participants. The first two sessions occurred back to back on March 30th and March 31st (nine participants total). After these sessions, the results from the completed questionnaires and focus group discussions were reviewed, and it was

determined that the completed problem lists would not be usable (because of the bug in the software, and because some students had not had time to complete their problem lists). At this point the developers saw that there were some questions left unanswered by the questionnaire, and that the lack of problem list data also left some questions unanswered. To compensate for this deficit, some of the items on the questionnaire were rewritten, some items were added, and the rewritten questionnaire was used for the last two small group sessions (the last 11 participants). Also, time taken to complete the problem lists was recorded for the last two small group sessions. The last two small group sessions occurred on April 12th and 13th. The specific factors measured will be reported for these two sets of small group sessions in terms of clarity, feasibility, and impact. The questionnaire used for the first two small group sessions can be found in Appendix E. The adapted questionnaire used for the second two small group sessions can be found in Appendix F.

Clarity

The first two small group sessions.

Many of the items in the questionnaire were used to indicate clarity. These included the items extracted from Shneiderman (1992) which measure general user reactions to the program as well as more specific factors such as to what extent the display layouts make tasks simpler. Many comments relevant to clarity also emerged in the comments for items 14-17.

The focus group sessions also provided information concerning the clarity of the PLG. Responses to general questions such as “Could you tell what to do next?” provided insight into clarity.

The notes taken by the observer during the small group session also informed the clarity question. These, in fact were some of the best sources of information about usability problems because the observer was able to see what was occurring first hand.

The last two small group sessions.

While some of the items in the questionnaire relating to clarity were reworded somewhat for the last two small group sessions, the indicators of clarity were not appreciably different for these groups than they were for the previous two.

Feasibility

The first two small group sessions.

The questionnaire was also used to indicate feasibility. Specifically items 8-13 were intended to gauge the students’ general acceptance of the PLG and the context in which it was to be used. The focus group and observation data were also useful indicators of feasibility.

The last two small group sessions.

In addition to the indicators used in the first two small group sessions, several indicators of feasibility were added to the questionnaire for the second set of small group sessions. These included items 7, and 10 which had to do with whether or not students felt they could grow accustomed to using the PLG for all their cases, and to what extent they were bothered by technical problems. Time taken to complete the cases was also record for these sessions, another indicator of feasibility.

Impact

The first two small group sessions.

Items 12 and 13 on the questionnaire were intended to measure the students' perception of the impact of the PLG. (These items sought information regarding whether the students felt that the PLG made organizing their thoughts and understanding clinical pathology harder or easier.) Unfortunately, the more direct indicator of impact -- the students' completed problem lists, were not available.

The last two small group sessions.

Two items intended to measure impact were added to the questionnaire for the second set of small group sessions (items 6a and 6b). These items sought the students' opinions of whether or not the PLG made their lists a) account for more lab data than normal and b) more or less precise than normal. The bug which would have made the completed problem lists available for analysis was still not completely fixed at the time the second two small group sessions were conducted.

Small Group Evaluation -- Results

General Results

The data gathered and analyzed from the small group evaluations suggest the following:

1. The problem list generator is considered usable by the students (in terms of clarity).
2. The problem list generator is considered somewhat usable by the students in terms of feasibility.
3. The students, on average, do not feel that the problem list generator will have any substantial impact on their comprehension of Clinical Pathology.
4. Most students enjoy the convenience of being able to hand in their cases automatically, and most students enjoy the availability of an expert list.
5. Students are divided in their acceptance of computer use in general.
6. Students are divided in their acceptance of the structure imposed by the problem list generator.

Results from the Analysis of the Questionnaire

Results of the analysis of data gleaned from the questionnaire given to the first 2 sets of small group participants (9 total) are summarized in Table 3.1. This questionnaire can be found in its entirety in Appendix E. As can be seen, students were relatively positive about their ability to use the PLG, but their overall acceptance of it as a viable alternative to what they are currently doing is somewhat mixed. Furthermore, most students don't agree that using the PLG will make organizing their thoughts about a case any easier, but on average the students did feel that using the PLG would improve their understanding of Clinical Pathology.

To make these results more easily interpretable, specific items were selected to represent Clarity, Feasibility, and Impact. The right most column of Table 3.1 indicates which construct each item was used to indicate. For example, the items used to indicate "Clarity" were items 1-6. The questionnaire being referred to can be found in Appendix E. The items used to indicate "Clarity" were meant to measure how easy it was for students to tell what to do with the software, and to give some overall indication of "usability". The items used to indicate "Feasibility" were meant to measure students' general acceptance of and ease of use of the software compared to what they do now. The items used to measure "Impact" asked the students to indicate whether they felt the PLG would make it easier or harder to a) organize their thoughts, and b) understand Clinical Pathology. All items in the questionnaire were evaluated on an eleven point scale (0-10), as taken from Shneiderman (1992). The reliability estimate for the construct of Clarity was not particularly high for this group ($\alpha = .36$). Items 1, 2, and 6 correlated well, as did items items 3, 4, and 5. However, these two sets of items did not correlate well together. The reliability measure for feasibility was quite good ($\alpha = .87$). The measures for impact were relatively reliable ($\alpha = .58$). Mean scores for these constructs can be found in Table 3.2. As can be seen in Table 3.2, on average, both the clarity and feasibility items scored about halfway between a total unconditional agreement, and neutrality, and, even the minimum items were higher than neutral, indicating that most students agreed, in general that the PLG was usable, in terms of clarity and feasibility. The average Impact estimate seems to indicate that these students didn't feel like the PLG would have much affect, overall, on their learning.

Table 3.1

Means, standard deviation, median, maximum, minimum and construct assignment by item for Small Group Evaluation sessions 1 and 2 (From the questionnaire in Appendix E)

Questionnaire Item:			<u>M ± SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
1. Operations relate to tasks	distantly	0 1 2 3 4 5 6 7 8 9 10	7.8 ± 2.0	9	9	3	9	Clarity
		closely NA						
2. Display layouts make tasks simpler	never	0 1 2 3 4 5 6 7 8 9 10	7.0 ± 2.5	8	9	1	9	Clarity
		always NA						
3. Sequence of displays	confusing	0 1 2 3 4 5 6 7 8 9 10	7.8 ± .83	8	9	6	9	Clarity
		clear NA						
4. Learning the Operation	difficult	0 1 2 3 4 5 6 7 8 9 10	7.9 ± 1.4	8.5	9	6	8	Clarity
		easy NA						
5. Human memory limitations	overwhelmed	0 1 2 3 4 5 6 7 8 9 10	9.4 ± .73	10	10	8	9	Clarity/ Feasibility
		are respected NA						
6. Overall navigation (Getting around in the program)	irritating	0 1 2 3 4 5 6 7 8 9 10	7.2 ± 1.6	8	9	5	9	Clarity
		easy NA						

Table 3.1. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
7(a). Overall reactions	terrible	wonderful	7.1 ± 1.6	8	8	3	9	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(7b)	frustrating	satisfying	6.7 ± 2.3	8	9	2	9	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(7c)	uninteresting	interesting	7.4 ± 1.8	8	10	5	9	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(7d)	dull	stimulating	7.0 ± 1.6	7	10	5	9	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(7e)	difficult	easy	7.3 ± 1.2	8	9	5	9	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						

Table 3.1. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
8. Doing my problem list homework using the problem list generator will be easier than the way I do it now.	definitely not 0 1 2 3 4 5 6 7 8 9 10	absolutely NA	5.3 ± 2.0	5	9	3	9	Feasibility
9. I like being able to do my problem lists on a computer.	definitely not 0 1 2 3 4 5 6 7 8 9 10	absolutely NA	6.7 ± 2.8	7	10	1	9	Feasibility
10. I like being able to do my problem lists on the web.	definitely not 0 1 2 3 4 5 6 7 8 9 10	absolutely NA	6.9 ± 2.7	7	10	2	9	Feasibility
11. I like having my problem lists turned in automatically as soon as I finish them.	definitely not 0 1 2 3 4 5 6 7 8 9 10	absolutely NA	8.6 ± 2.4	10	10	4	9	Feasibility

Table 3.1. (continued)

Questionnaire Item:			<u>M ± SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
12. Using the problem list generator makes it easier for me to organize my thoughts about a case.	definitely not	absolutely	4.6 ± 2.5	5	9	0	9	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						
13. Using the problem list generator will make understanding clinical pathology. . .	harder	easier	6.5 ± 1.2	6.5	8	5	8	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						

Table 3.2

Clarity, Feasibility, and Impact construct means for Small Group Evaluation sessions 1 and 2(From the questionnaire in Appendix E)

Construct Measured	Mean	Median	Minimum	Maximum	Count
Clarity	7.86	7.83	6.17	8.67	9
Clarity	7.86	7.83	6.17	8.67	9
Feasibility	7.24	7.70	5.10	8.60	9
Feasibility	7.24	7.70	5.10	8.60	9
Impact	5.17	5.00	.00	8.50	9
Impact	5.17	5.00	.00	8.50	9

Note. The items from the questionnaire used to indicate Clarity are items 1-6. Items 5, 7a-7e, and 8-11 were used to indicate Feasibility. Items 12-13 were used to indicate Impact.

Table 3.3 accomplishes the same purpose as did Table 3.1, but for the second two sets of small group participants. The complete questionnaire used for these items can be found in Appendix F, and was re-written in hopes of improving the precision with which Clarity, Feasibility, and Impact could be assessed. While lower than the results of the first two small group sessions, these results continue to indicate general agreement from most students that the PLG is usable in terms of clarity and feasibility. As with the first groups, most felt that the PLG was likely to have little impact on learning, or the quality of their problem lists.

As with the first small group evaluation, items were divided into the constructs of Clarity, Feasibility, and Impact as a tool for more easily communicating the results. And, as before, the construct to which each item was applied can be found in the right hand column. The reliability estimates for the constructs of Clarity, Feasibility, and Impact, ($\alpha = .42, .85,$ and $.56$ respectively) either improved somewhat or remained stable. As will be seen with the field trial, the relatively low alpha values for Clarity and Impact might be due somewhat to the small sample size, as they improve markedly for the same questionnaire with a larger group. Table 3.4, (as Table 3.2) communicates the overall questionnaire results in terms of each construct.

Table 3.3

Means, standard deviation, median, maximum, minimum and construct assignment by item for Small Group Evaluation sessions 3 and 4 (From the questionnaire in Appendix F)

Questionnaire Item:			<u>M ± SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
1. Using the Problem List Generator made me feel. . .	overwhelmed	comfortable	5.9 ± 1.6	6	8	3	11	N/A ^a
	0 1 2 3 4 5 6 7 8 9 10	NA						
2. Display layouts make tasks simpler	never	always	6.8 ± 2.4	7.5	10	2	11	Clarity
	0 1 2 3 4 5 6 7 8 9 10	NA						
3. Sequence of displays	confusing	clear	6.6 ± 3.0	7	10	0	11	Clarity
	0 1 2 3 4 5 6 7 8 9 10	NA						
4. Learning How to Use It	difficult	easy	6.7 ± 2.6	8	10	2	11	Clarity
	0 1 2 3 4 5 6 7 8 9 10	NA						
5. Compared to how I do problem lists now, using the Problem List Generator to do cases is. . .	harder	easier	4.6 ± 3	4	10	1	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						

Table 3.3. (continued)

Questionnaire Item:		<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct	
6. Using the Problem List Generator. . .								
6 (a) Made me account for more lab data than I normally do	No/less 0 1 2 3 4 5 6 7 8 9 10	Yes/More NA	5.3 ± 3.0	5	9	0	11	Impact
6 (b) Made my problem list more precise	No/less 0 1 2 3 4 5 6 7 8 9 10	Yes/More NA	4 ± 2.9	4	9	0	11	Impact
6 (c) Made me remember too much	Yes/Too much 0 1 2 3 4 5 6 7 8 9 10	No NA	7.8 ± 2.9	10	10	3	11	Clarity
7. I could get used to using the Problem List Generator to do all my cases	No Way 0 1 2 3 4 5 6 7 8 9 10	Definitely NA	6.9 ± 3.0	7	10	2	11	Feasibility
8. Overall navigation (Getting around in the program)	irritating 0 1 2 3 4 5 6 7 8 9 10	easy NA	5.8 ± 2.6	5	9	2	11	Clarity
9. Overall reactions	terrible 0 1 2 3 4 5 6 7 8 9 10	wonderful NA	6.3 ± 1.6	7	8	3	11	Feasibility

Table 3.3. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
(9b)	frustrating	satisfying	6.1 ± 2.3	6	10	3	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9c)	uninteresting	interesting	6.6 ± 2.0	7	10	4	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9d)	dull	stimulating	6.6 ± 2.2	7.5	10	3	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9e)	difficult	easy	5.8 ± 2.0	5	8	3	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
10. I was frustrated by technical problems	Yes	No	4.9 ± 3.0	5	10	0	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
11. Doing my problem list homework using the problem list generator will be easier than the way I do it now.	definitely not	absolutely	4.6 ± 2.7	5	10	0	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						

Table 3.3. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
12. I like being able to do my problem lists on a computer.	definitely not	absolutely	5.6 ± 3.1	5	10	0	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
13. I like being able to do my problem lists on the web.	definitely not	absolutely	5.6 ± 3.0	5	10	0	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
14. I like having my problem lists turned in automatically as soon as I finish them.	definitely not	absolutely	7.5 ± 3.3	9	10	0	11	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
15. Using the problem list generator makes it easier for me to organize my thoughts about a case.	definitely not	absolutely	4.6 ± 3.5	4	10	0	11	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						
16. Using the problem list generator will make understanding clinical pathology. . .	harder	easier	5.6 ± 2.3	5	10	3	11	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						

^aItem one was not used in the analysis because a large number of respondents changed the word “overwhelming” to something else, like “irritating” or “annoying.”

Table 3.4

Small Group Evaluation sessions 3 and 4 (From the questionnaire in Appendix F)

Construct	Mean	Median	Maximum	Minimum	Count
Clarity	6.75	7.20	9.00	4.60	11
Clarity	6.75	7.20	9.00	4.60	11
Feasibility	5.92	5.50	9.25	4.33	11
Feasibility	5.92	5.50	9.25	4.33	11
Impact	4.83	4.50	9.25	2.50	11
Impact	4.83	4.50	9.25	2.50	11

Note. Items 2-4, 6c and 8 were used as indicators of Clarity, items 5, 7, 9a-9e, and 10-14 were used as indicators of feasibility, and items 6a, 6b, 15, and 16 were used as indicators of impact.

Table 3.5

Mean of all small group sessions

Construct	Mean	Count
Clarity	7.3	20
Clarity	7.3	20
Feasibility	6.58	20
Feasibility	6.58	20
Impact	5	20
Impact	5	20

Table 3.5 represents the mean scores across all 20 participants for the constructs of Clarity, Feasibility, and Impact, as measured by the two questionnaires, and it is consistent with the results discussed to this point (as can be expected.) In general, the participants in the small group sessions found the software usable in terms of

clarity (usability) and feasibility, but on the average, they didn't feel that using the PLG would have much effect on their problem lists or overall learning.

Results from observation, student comments, and the focus group conversations

Comments from students emerged through open-ended questions on the questionnaires, remarks made while using the problem list generator, and remarks made during focus group conversations. These comments can generally be broken into four categories: a) things the students liked about the problem list generator, b) things the students didn't like about the problem list generator, c) things that were widely disagreed upon, and d) reasons the students gave for dislikes.

What they liked

Most students liked the fact that the PLG makes it possible to hand in problem lists automatically. Students also liked the fact that they could see the expert problem list and find out "how they did". Some organizational features of the PLG, such as the fact that used data abnormalities are "checked off" and the fact that the data abnormalities only have to be typed once were also very popular. Most students also liked the drag and drop feature.

What they didn't like

Most students were frustrated by bugs, technical glitches, and unresolved interface problems. For example, while waiting for the network to transfer data, the PLG frequently appears "crashed" for periods of up to 30 seconds or more. In one student's words, "I'm not particularly computer literate and when glitches pop up - I get very frustrated." Many students also mentioned that the PLG did not allow them to see the "big picture" as they could on paper (i.e. not all data abnormalities, nor the whole problem list are present to the eye at once.)

Items of disagreement

Students were quite divided in their opinions of the control exerted by the PLG. Many liked the fact that the PLG made them spell words correctly, or would not allow them to proceed to a problem list until they had identified all the data abnormalities. These features of the PLG were frustrating to other students who wished to have the freedom to make mistakes. Still other students wanted the PLG to provide immediate answers (i.e. correct spelling) rather than providing feedback and opportunities to retry. Many students were very excited about the fact that the PLG was computer based and web based. Others didn't like the PLG because it was computer based and web based.

General reasons the students gave for dislikes

Some of the students' reasons for their dislikes had nothing to do with the PLG itself. For example, many of the students indicated that they found the PLG harder or less desirable than their current method of completing and turning in cases because it was different from the way they were accustomed to doing their lists. These students indicated that they would like the PLG better than their current method of completing cases had they learned to create problem lists with the PLG. Some students also indicated a general lack of familiarity with computers as a reason for their discomfort with the PLG. Other students were afraid that it would be harder to collaborate with friends making problem lists, and that the limited number and locations of computers would limit the times and places that problem lists could be completed.

Field Trial - Methods

Introductory Note

The original field trial was conducted with 24 participants who were asked to complete one case using the PLG on their own time and using any computer they chose. This occurred in the final two weeks of the semester, and only about half (11) of the field trial volunteers had completed a case by the time the class was over. Most of the remaining participants all completed cases using the PLG after the semester had ended. (Three participants weren't able to finish problem lists at all because they either were unable to figure out how the PLG worked, or because they encountered technical difficulty). This meant that about half of the participants did cases that they had seen and worked on paper previous to doing them with the PLG. Therefore, while the results from these participants proved useful for measuring clarity and feasibility, the problem lists generated were not useful for measuring impact. Approximately three to five weeks after the end of the semester, 10 students who had just completed the course were engaged to complete 2 new cases one using the PLG and one on paper. There are, then, 3 groups of field trial participants: those who participated in the original field trial and completed a case that was new to them (Group A), those who participated in the original field trial and completed a case they had seen before (Group B) and, those who completed both a new PLG-based case and a new paper-based case three to five weeks after the semester ended (Group C).

Participants

All participants in the field trial are current students at the Virginia Maryland Regional College of Veterinary Medicine, and took the Clinical Pathology course (VM 8414) in Spring semester of 1999. All are between 20 and 40 years of age with most being in their early twenties. Of all the participants, 12 are male and 22 are female. Those in Groups A and B are unpaid volunteers, but received extra credit in VM 8414 for participating. Those in Group C were paid \$30.00 for participating in the study and were self-selected from the group of veterinary students in their class still remaining in town after the semester had ended.

Throughout the formative evaluation process, some care was taken to ensure that advanced, medium, and below average learners (in terms of ability) were all well represented in the evaluation process. It was hoped that the PLG would be particularly useful to those learners who were below average in their ability with Clinical Pathology, and that it would not have adverse effects on students of any ability level. To determine to what extent there were ability-predicted differences in the student outcomes with the PLG, students in Groups A and C were further divided into three groups, according to ability as indicated by their final exam score. (The final exam occurred between one and three weeks after Group A participated in the field trial, and approximately 3 to 4 weeks before Group C participated in the field trial.) The students classified as high ability (approximately 33% of the class) scored 90% or higher on the final exam. The students classified as average ability (approximately 33% of the class) scored between 80% and 90% on the final, and those classified as low ability (approximately 33% of the class) scored 78% or lower on the final exam.

Group A

There were 11 students in the original field trial who completed new problem lists. Of these, eight were female and three were male, 5 were considered low ability, 3 medium ability, and 3 high ability. Three additional members of this group used the PLG and completed questionnaires, but were unable to complete a problem list due to technical problems or an inability to understand the interface. Of these, two were females and one was a male, one was high in ability, one average, and one low.

Group B

There were ten students in the original field trial who completed a case they had already seen. Of these, eight were female and two were male. Four were considered low in ability, four were considered of average ability, one was considered high in ability, and ability information was not available for the other.

Group C

Ten students completed a new case with the PLG and a new case on paper. Of these 5 were female and 5 were male. Two were considered low ability, 4 were considered average ability, and 4 were considered high in ability.

Procedures

Procedures for Students in Groups A and B

The twenty four students in groups A and B were asked to complete at least one case using the problem list generator. Since the purpose of a field trial is to evaluate the instruction in a real setting (Tessmer, 1993), the students were not required to use any specific location for completing their cases. Some completed their cases at home, and

some at one of the several computer labs at the College of Veterinary Medicine. A “help” section was provided for the students at the course website (a link to it was available on the page from which the PLG was started). This help section explains how to use the PLG, and describes bugs and work-arounds. Just as would be the case in an authentic implementation situation, students were permitted to seek help from other students, faculty, and staff. While completing the field trial, students were asked to record how long it took to complete each case using the PLG.

Procedures for students in Group C

Students in Group C completed their problem lists, both the paper lists, and the PLG-based lists, in the Medical Informatics Lab -- a small lab housing 4 work stations. The participants were asked to complete two cases (which they had not seen previously), one using the PLG, and one as they were accustomed to doing on paper. Half of the students (by random assignment) used the Problem List Generator first, and the other half did a case on paper first. Also, each case was done by half of the students using the PLG and by half of them on paper (again, by random assignment.) This resulted in each student completing two lists -- one using the PLG, and one on paper. Students recorded how long it took to complete each case.

Students in Group C were offered help as needed while working on the PLG in the lab. They were also informed of bugs and work-arounds. While all of the information they received was available to students in Groups A and B as well, they had much more ready access to the information. Therefore, information from students in group C offers a less realistic view of clarity and feasibility than does the information for students in groups A and B. However, the impact information (the completed problem list) is comparable.

Data Sources

There were two main sources of data for the field trial: the questionnaires and the students' completed problem lists. Because students completed this exercise with a number of computers at a number of locations, it was not feasible to collect observational data. However, participants did supply a great deal of “qualitative” information in the open answer sections of the questionnaire. Students also reported how much time was taken to complete each case. As was the case with the other evaluation efforts of the study, the purpose of the field trial was to assess clarity, feasibility, and impact, with an emphasis on impact.

Clarity

The same questionnaire that was used for the second set of small group evaluations was used for the field trial, and the same questions used to indicate clarity for the second set of small groups (items 2-4, 6c, and 8) provided an indicator for clarity for the field trial.

Feasibility

The questionnaire was also used to indicate feasibility. Specifically items 5, 7, 9a-9e, and 10-14 were intended to gauge the students' general acceptance of the PLG and the context in which it was to be used.

Impact

Items 6a, 6b, 15 and 16 on the questionnaire were intended to measure the students' perception of the impact of the PLG. However, completed problem lists were considered the primary indicators of impact. The nature of the problem lists, and how they were evaluated will be described below.

As explained in chapter 2, a problem list is a list in outline format that communicates causal relationships between items of data (including mechanisms, data abnormalities, and "other" (including physical exam findings, signalment, and history information)). Causal relationships are communicated in problem lists by placing causal items above and to the left of affected items. There can be multiple nested levels of causal relationships.

The problem lists were evaluated as follows.

1. Professional typists were engaged to type up the paper-based problem lists. These typed-up lists were labeled in such a way as to maintain the confidentiality of the participants.
2. Each case was "solved" by an expert in Clinical Pathology (the professor for the course). This solution resulted in an "expert problem list."
3. The professor identified critical or highly important items which should be present in all cases.
4. A veterinary student who had completed the clinical pathology course analyzed all of the problem lists (student and expert lists). For this analysis, the student identified a) all causal relationships communicated in each problem list, b) all data abnormalities present in each problem list, c) all mechanisms present in each problem list, d) all "high priority items" present in each problem list and e) all "other" data (signalment, history, and physical exam findings) present in each problem list.
5. The same veterinary student compared each student problem list to the corresponding expert problem list and identified, a) the percentage of causal relationships, data abnormalities, mechanisms, high priority items and "other" identified in each learner list that were also present in the expert list, b) the number of causal relationships, data abnormalities, mechanisms, high priority items, and "other" identified in each learner list that were not present in the expert list, c) the number of causal relationships, data abnormalities, mechanisms, high priority items, and "other" identified in each learner list that were incorrect (frequently the same as (b)), and d) whether or not the diagnosis in the learner list was correct.

The expert (professor) corroborated the evaluator's analysis by providing consultation to the student on a regular basis, and by reviewing approximately 25% of the analyzed cases (randomly selected). The professor's evaluation of those cases matched the evaluator's evaluation well. Across all cases compared, there was a 99.2% agreement on Causal Relationships, a 99.4% agreement on Data Abnormalities, a 99.1% agreement on Mechanisms, a 97.7% agreement on High Priority items, and a 100% agreement on the "other" information.

Field Trial Evaluation -- Results

General Results

The data gathered and analyzed from the field trial evaluation suggests the following:

1. Based on comparisons to expert problem lists, there is no statistically significant evidence to suggest that minimal exposure to the PLG will affect the quality of problem lists produced by learners who are used to other ways of analyzing and synthesizing laboratory data. Trends, however, do seem to favor the PLG.
2. There is no statistically significant evidence to suggest that the PLG favors any particular group of learners in terms of ability level. Trends, however, seem to suggest that the PLG might have a positive impact on the completeness and precision of problem lists for low and average ability learners.
3. There is evidence to suggest that the PLG, in its present state, is not intuitive enough to be used by all students without some introductory explanation.
4. There is evidence to suggest that fixing currently known bugs and usability problems in the PLG will favorably affect its clarity and feasibility.

Specific Findings

As discussed in the methods section, there were three subgroups involved in the field trial, designated for the purposes of this document as groups A, B, and C. Group A was comprised of eleven students who worked with new cases and were part of the original "authentic" field trial. Group B was comprised of ten students who were involved in the original field trial and who completed cases they had already seen. Group C was comprised of 10 students who completed a new case on the PLG and a different new case on paper in the Medical Informatics Lab several weeks after the semester had ended. All participants had been students in the same Clinical Pathology course in the Spring of 1999. The problem lists from Groups A and C were considered usable for measuring impact, because all of these students used the PLG for completing cases to which they had not been exposed. The data from the questionnaires for Groups A and B will be used for indications of clarity and feasibility because these students had to deal with the PLG in an authentic field trial setting in which they were "on their own." The

questionnaire data from Group C will not be used to indicate clarity or feasibility because these students had more ready access to help while using the PLG.

Impact

The completed problem lists were the primary source of information regarding impact. Reliable data from the problem lists was available for 21 participants (Groups A and C). The problem lists completed with the PLG were compared to the problem lists completed on paper for each participant using an analysis of variance (ANOVA). Means were also compared by ability level. Because the problem lists chosen by the students were not all equivalent in difficulty, a difficulty index was used as a covariate to control for the difficulty of the cases. The difficulty index (DI) was arrived at for each case by summing, from the expert list, the number of causal relationships, data abnormalities, mechanisms, and clinical signs present. The comparative means are presented in Table 3.6 for the entire group as well as for the three ability groups.

These means reveal some trends, most of which appear to favor the PLG. Overall, for example, more causal relationships, mechanisms and data abnormalities were identified using the PLG than the paper-based method, and fewer erroneous data abnormalities and causal relationships were identified using the PLG than using the paper-based method. Of all overall differences due to tool, only two approached significance, however (and one only remotely): percent of causal relationships identified correctly $F(1, 35) = .8, p = .38$ and percent of mechanisms identified correctly $F(1, 35) = 2.7, p = .11$. Both of these differences favored the PLG.

The means representing tool by ability seem to favor the PLG for average and below average achievers over most measures (a notable exception being the value for

Table 3.6
Overall Means for PLG and Paper and Means for PLG and Paper by Ability Level Adjusted for Difficulty of Case

Tool	CAUSAL ^a	CAUSALW ^b	DA ^c	DAW ^d	MECH ^e	MECHW ^f	OTHER ^g	HIGH PRIORITY ^h	n
Overall:									
PLG	37 ± 3.4	3 ± .86	80 ± 3.8	.4 ± .17	49 ± 4.3	.8 ± .2	37 ± 5.2	73 ± 5.2	21
Paper	33 ± 3.4	4 ± .86	80 ± 3.8	.5 ± .17	39 ± 4.3	.7 ± .2	44 ± 5.2	67 ± 5.2	21
High Ability									
PLG	35 ± 5.9	4 ± 1.5	73 ± 6.7	.8 ± .30	50 ± 7.5	.8 ± .38	43 ± 9.0	80 ± 9.0	7
Paper	38 ± 5.8	4 ± 1.5	85 ± 6.7	.6 ± .30	47 ± 7.5	1 ± .37	48 ± 9.0	74 ± 9.0	7
Medium Ability									
PLG	38 ± 5.9	2 ± 1.5	85 ± 6.7	.6 ± .30	51 ± 7.5	.7 ± .38	27 ± 9.0	77 ± 9.0	7
Paper	31 ± 5.9	3 ± 1.5	75 ± 6.6	.3 ± .30	35 ± 7.5	.6 ± .37	54 ± 9.0	64 ± 9.0	7
Low Ability									
PLG	37 ± 5.9	4 ± 1.5	81 ± 6.7	.5 ± .30	45 ± 7.5	.9 ± .37	42 ± 9.0	62 ± 9.0	7
Paper	29 ± 5.9	5 ± 1.5	79 ± 6.7	.7 ± .30	35 ± 7.5	.6 ± .37	29 ± 9.0	63 ± 9.0	7

Note. All means are given ± the Std. Error.

^a % of causal relationships correct

^b Number of causal relationships incorrect

^c % of data abnormalities correct

^d Number of data abnormalities incorrect

^e % of mechanisms correct

^f Number of mechanisms incorrect

^g % of other information correct.

^h% of high priority items identified correctly

Table 3.7

Analysis of Variance for use of PLG Adjusted for Difficulty of Case

Source	df	<u>F</u>							
		Between Subjects							
		Causal	Causal Wrong	Data Abns.	Data Abns. Wrong	Mechanisms	Mechanisms Wrong	Other	High Priority
Tool (PLG or Paper)	1	.80	.17	.00	.24	2.7	.05	.81	.63
Ability	2	.18	1.2	.02	2.0	.73	.26	.64	1.3
Tool × Ability	2	.52	.28	1.4	.42	.35	.30	2.6	.26
Error	35								

Note. None of the F values were significant with an alpha level of .05

“other items” for average learners). For above average learners the tool seems to have had little effect for most measures, and noticeable differences seem to favor the more familiar paper-based method. The interaction differences do approach significance for DA and Other, ($F(2, 35) = 1.4, p = .27$ and $F(2, 35) = 2.6, p = .09$ respectively).

The mean comparisons are encouraging, and certainly suggest that further testing with a larger sample size is in order. However, as seen in Table 3.7, none of the differences tested (overall or interactions) were significant at the $p = .05$ level using an analysis of variance.

The statistical findings are generally consistent with the participants' own opinions of how much the PLG would affect the accuracy and completeness of their problem lists. As seen in Tables 3.8 and 3.9, when presented with the same questionnaire that was given to the small group participants, the field trial participants felt that using the PLG was neither likely to improve or damage the quality of their problem lists.

Clarity and Feasibility

Table 3.8 reveals overall mean responses by item to the questionnaire (as seen in Appendix F) for participants in the field trial, and as before, indicates which items were used as indicators of which constructs. (Recall that there are fewer values reported for the feasibility and clarity scores than for the impact scores because only scores for Groups A and B -- those who had an “authentic” field trial experience -- were used as Clarity and Feasibility indicators.) Most item means fell above the midpoint, though the Clarity and Feasibility measures were somewhat lower, overall, than the corresponding small group measures had been. (This is to be expected, as students were left to struggle on their own in the Field Trial more than in the Small Group sessions.) The impact scores were roughly as they had been with the small group evaluations, though, interestingly the mean for item 16 was a relatively high $6.2 (\pm 1.8 \text{ SD})$, suggesting that, in spite of the weaknesses they found in the PLG, students did feel, on the average, that it would somewhat improve their understanding of Clinical Pathology.

The same items that were used as indicators of Clarity, Feasibility, and Impact for the second set of Small Group sessions were again used for indicating those constructs with the field trial results. With a larger sample size, the overall reliability estimates improved overall with Clarity, Feasibility, and Impact alpha values of $\alpha = .75$, $\alpha = .86$ and $\alpha = .78$ respectively, suggesting that the items correlate well, and that previous low alphas were due to the small sample sizes in the small group evaluation sessions.

Table 3.8

Means, standard deviation, median, maximum, minimum and construct assignment by item for the Field Trial Evaluation Sessions

(From the questionnaire in Appendix F)

.Questionnaire Item:		<u>M ± SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
1. Using the Problem List Generator made me feel. . .	overwhelmed 0 1 2 3 4 5 6 7 8 9 10 comfortable NA	5.9 ± 2.4	7	9	0	34	N/A ^a
2. Display layouts make tasks simpler	never 0 1 2 3 4 5 6 7 8 9 10 always NA	6.7 ± 2.1	7	9	0	24	Clarity
3. Sequence of displays	confusing 0 1 2 3 4 5 6 7 8 9 10 clear NA	6.8 ± 2.7	8	10	2	24	Clarity
4. Learning How to Use It	difficult 0 1 2 3 4 5 6 7 8 9 10 easy NA	5.0 ± 2.4	5	9	0	24	Clarity
5. Compared to how I do problem lists now, using the Problem List Generator to do cases is. . .	harder 0 1 2 3 4 5 6 7 8 9 10 easier NA	4.5 ± 2.8	4.0	9	0	24	Feasibility

Table 3.8. (continued)

Questionnaire Item:		<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
6. Using the Problem List Generator. . .		5.5 ± 3.2	5	10	0	34	Impact
6 (a) Made me account for more lab data than I normally do	No/less 0 1 2 3 4 5 6 7 8 9 10 Yes/More NA						
6 (b) Made my problem list more precise	No/less 0 1 2 3 4 5 6 7 8 9 10 Yes/More NA	5.8 ± 2.6	6	10	0	34	Impact
6 (c) Made me remember too much	Yes/Too much 0 1 2 3 4 5 6 7 8 9 10 No NA	7.1 ± 2.6	8	10	0	24	Clarity
7. I could get used to using the Problem List Generator to do all my cases	No Way 0 1 2 3 4 5 6 7 8 9 10 Definitely NA	6.5 ± 3.2	7	10	0	24	Feasibility
8. Overall navigation (Getting around in the program)	irritating 0 1 2 3 4 5 6 7 8 9 10 easy NA	4.8 ± 2.5	5	9	0	24	Clarity
9. Overall reactions	terrible 0 1 2 3 4 5 6 7 8 9 10 wonderful NA	5.7 ± 2.2	6	9	0	24	Feasibility

Table 3.8. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
(9b)	frustrating	satisfying	4.4 ± 2.6	4	9	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9c)	uninteresting	interesting	7.4 ± 1.9	8	10	3	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9d)	dull	stimulating	7.2 ± 1.7	8	10	3	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
(9e)	difficult	easy	5.3 ± 2.5	5	9	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
10. I was frustrated by technical problems	Yes	No	7 ± 2.9	8	10	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
11. Doing my problem list homework using the problem list generator will be easier than the way I do it now.	definitely not	absolutely	4.6 ± 2.4	5	10	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						

Table 3.8. (continued)

Questionnaire Item:			<u>M</u> ± <u>SD</u>	<u>Mdn</u>	Max.	Min.	n	Construct
12. I like being able to do my problem lists on a computer.	definitely not	absolutely	6.4 ± 3.0	6	10	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
13. I like being able to do my problem lists on the web.	definitely not	absolutely	6.1 ± 3.0	6	10	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
14. I like having my problem lists turned in automatically as soon as I finish them.	definitely not	absolutely	8.1 ± 3.0	9.5	10	0	24	Feasibility
	0 1 2 3 4 5 6 7 8 9 10	NA						
15. Using the problem list generator makes it easier for me to organize my thoughts about a case.	definitely not	absolutely	5.2 ± 2.5	5	10	0	34	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						
16. Using the problem list generator will make understanding clinical pathology. . .	harder	easier	6.2 ± 1.8	5	10	3	34	Impact
	0 1 2 3 4 5 6 7 8 9 10	NA						

^aItem one was not used in the analysis because a large number of respondents changed the word “overwhelming” to something else, like “irritated” or “annoyed.”

Table 3.9 reports the means for each construct for the Field Trial evaluation. Overall, it appears that the participants in the field trial found the PLG more usable than not, in terms of Clarity and Feasibility (6.04 for clarity and 6.03 for feasibility on a scale of 0 to 11). They also, generally, felt that the PLG was unlikely to affect overall learning or problem list quality as measured by the impact construct which had a mean of 5.6.

Table 3.9

Questionnaire results for the students participating in the field trial for the constructs of clarity, feasibility, and impact.

Construct	Mean	Median	Maximum	Minimum	Count
Clarity	6.04	6.40	8.80	2.75	24
Feasibility	6.03	6.09	8.91	1.11	24
Feasibility	6.03	6.09	8.91	1.11	24
Impact	5.61	5.75	9.25	.00	34
Impact	5.61	5.75	9.25	.00	34

Note. The items from the questionnaire (Appendix F) used to indicate Clarity are items 2-4, 6c, and 8. The items used to indicate feasibility are items 5, 7, 9a-9e, and 10-14, and the items used to indicate impact are items 6a, 6b, 15, and 16.

While the questionnaire findings are useful, they leave many of the “why’s” unanswered. And these are, in fact, the most important questions to answer in a formative evaluation. Fortunately, many reasons that the students scored the questionnaire as they did are identifiable in the comments provided by the participants in the fairly open-ended questions at the end of each questionnaire. Many of these factors will only make sense to those who have read the interface design (Appendix A). But as these details are essential to an understanding of what the questionnaire data really says about the PLG, I will review them briefly here.

Factors that lowered clarity and feasibility ratings

In general what the students didn't like fell into six general categories: a) Features that are designed into the PLG, b) The fact that the PLG is different, c) Things which are unclear how to do, d) Things they would like the PLG to do that it doesn't do, e) Technical problems, and f) Contextual factors the developers can't control.

a) Characteristics present by design that one or more people did not like:

Some participants wanted features that would make the learning task less rigorous. For example, several participants wanted pick-lists of data abnormalities, mechanisms, and diagnoses, so that they would not have to generate them from memory

or from notes or textbooks. Most other problems with designed characteristics were matters of preference, and for everyone who didn't like a certain feature, there were one or two participants who specifically mentioned that they did. For example, one participant didn't like the notes fields, while many others used them extensively.

b) The fact that the PLG is new/different

As was the case with the field trial, many participants said that they would probably prefer the PLG to the way they currently do cases if they had used the PLG from the start.

c) Usability problems

Several functions were unclear. For example, it took several people awhile to figure out that you must type in the names of the DA's and hit enter before going on to the next one. Another person did not understand how to list clinical signs individually for the problem list. Several people had general problems figuring out why mechanisms are inserted automatically in the problem list, and several people had trouble moving from one screen to the next or figuring out at first what the main menu was for. Most of these issues were resolved after minimal use, but did present initial annoyances.

d) Desired additional features

Many participants requested additions to the PLG, which will not be enumerated here. Many of these changes had already been proposed and planned, but time and resources had not permitted them to be implemented at the time of the field trial. For example, an hourglass icon, for indicating system "wait time" has been planned, but not yet implemented.

e) Technical Problems

There were a variety of technical problems, the most notable of which involved frequent crashes when using Netscape for the PC, and a Mac-related problem in which the window in which the problem list is built will not show scroll bars when the amount of text outgrows the size of the window.

f) Other contextual factors

Three people mentioned that they aren't fond of working with computers in general.

Overall Observations

Despite the various problems described above, all but two of the participants were able to complete at least one case with the PLG with minimal help, and in that sense the PLG was demonstrably usable. It should be noted, however, that most participants did

need some explanation, particularly of how to work around bugs, to use the PLG effectively. This suggests that, before the PLG is implemented with any given group in its present state, the group should receive a brief description of how it works and how to avoid bugs. Overall, while the learners had a fairly long list of desired fixes, issues of clarity did not tend to impede their ability to successfully use the software with minimal direction.

Factors that improved the clarity and feasibility ratings

The learners also mentioned a variety of clarity/usability factors that they liked about the PLG. What they liked about the PLG can probably best be placed in three main categories: factors due to interface/instructional design, issues of convenience/context and general issues.

a) Factors due to design

A number of learners mentioned that the PLG made manipulating the data easier. For example, they appreciated only having to name the data abnormalities once, being able to drag and drop items, having used data abnormalities checked off as they were used, etc. Many participants liked the instant feedback of the expert problem list. Many also felt that the control mechanisms employed by the PLG (such as checking spelling and ensuring that data abnormalities were identified) made their lists more accurate and precise. In one participant's words, the PLG " makes you think about mechanisms -- not just indent to look pretty."

b) Convenience

Several participants felt that the PLG improved the convenience of creating problem lists. Several people specifically mentioned that they liked being able to turn the problem lists in on-line. Others made general positive references to "Automated features", saved "paper and writing time", the fact that the PLG is "on the computer", and the fact that the PLG is "on the web." (As evidenced in the previous section, these latter two features were not convenient for everyone.

c) General/hard to interpret positive comments

A number of users made general positive statements about the PLG, such as that it was "pretty cool", " a great idea in the works", "easier and more efficient", and "a handy tool."

In general, as clarity and feasibility ratings were both above average, the students' likes seem to slightly outweigh their dislikes, indicating a certain measure of success combined with a need for some improvement.

Time Data as an Indicator of Feasibility

Students in Group C completing the field trial recorded how long it took them to complete their cases using the PLG vs. how long it took them to complete their cases on paper. Data for both PLG cases and paper based cases was only available for Group C, making the sample size relatively small (10). Because only two, roughly equivalent cases were used with Group C, and because each case was assigned randomly to one of the tools (PLG or paper) for each participant, the difficulty index was not used in this comparison. An independent samples T-test was employed to compare the amount of time spent completing cases with the PLG and in the more familiar paper-based manner.

Table 3.10

Time Means Compared for the Field Trial

Tool	Mean	N	Std. Deviation	Std. Error Mean
Paper Problem List	27.5	10	4.7	1.5
PLG Problem List	54.9	10	11.2	3.5

As seen in Table 3.10, on the average, it took students about twice as long to complete a case using the PLG than it did using their accustomed paper-based method. This difference is significant, as seen in Table 3.11. Not too much should be made of this difference, however, as students were not at all familiar the PLG when they completed the field trial, and were, on the contrary, very familiar with their accustomed paper-based method for completing problem lists.

Table 3.11

Paired Samples T-test -- Time Taken to Complete the Problem Lists

	t	df	Sig. (2-tailed)	Mean Difference	Standard Error Difference
Time -- PLG vs. Paper	7.1	12.1	< .001	27.4	3.8

Note. Equal variances not assumed.

This chapter reported methods and results for the small group evaluations, and for the field trial. Chapter 4 will discuss the findings.

CHAPTER 4: DISCUSSION

The Best Laid Plans of Mice and Men . . .

When this evaluation was first planned, it was to include a half semester of full-scale implementation with the PLG. A time-series design was to be employed, comparing weekly quiz scores, confidence indicators, and randomly selected problem lists. Each month followed the previous one, with no end to bug fixes in sight, and it became clear that the project was going to take longer to implement than anticipated. As the end of the semester rapidly approached, (and as our 90 available subjects were gearing up for final exams and preparing to leave for the summer), the question became, not whether a good, solid, summative evaluation would be possible, but whether small group evaluations or a field trial would even be possible. Through many late nights, endless hours of debugging time, pleading for volunteer participants, and the generosity of professors and students, the small group evaluations and field trial were accomplished. Nonetheless, while this evaluation has been useful (its findings will be invaluable to the continued enhancement of the PLG), these findings are preliminary. I will discuss the findings using the same categories that have been used throughout the report -- impact, clarity, and feasibility.

Impact

The Problem Lists

The data provided by the problem lists provides no statistically compelling evidence that using the PLG makes any significant difference in the learners' ability to synthesize and classify clinical lab data. This might prove to be the final answer. However, there are good reasons to suspect that the trends that currently favor the PLG will prove statistically significant with more equal comparisons.

First, at the time the students completed the field trial, they had completed more paper-based lists than PLG-based lists at a ratio of anywhere between 84 and 93 to 1. The fact that a completely foreign way of dealing with lab data, including an unfamiliar medium (for some), a brand new interface/method, and a lack of the comfortable, familiar setting, resulted in problem lists that were just as good as the lists produced in the traditional fashion, is actually quite heartening.

Second, as laid out in chapter 1, there are compelling theoretical reasons that providing students with an expert scheme (such as the PLG) should have a positive effect on learning outcomes (particularly in the long term). This is consistent with the observation (though not statistically significant) that learners in the average and slow groups performed better with the PLG than they did using their traditional paper-based method. The same was not noted for advanced learners. This suggests that advanced learners had developed a scheme, throughout the course of the semester that allowed them to effectively deal with clinical laboratory data without a tool such as the PLG. Exposure to the PLG, then, either added nothing to the scheme which had been developed, or failed to support it (conflicted with it). Average and below average

learners, however, had not developed an effective scheme for dealing with laboratory data throughout the semester, so the PLG provided the framework that was missing. With sweat, time, money, and pride invested in the PLG, this is an attractive explanation to the researchers. And in spite of that, it does seem to be a reasonable proposition.

A number of bugs and technical problems plagued the implementation of the PLG during the field trial. It is possible that these technical problems affected the students' ability to focus on their lists.

The Questionnaire

It is interesting that students accurately predicted (overall) that using the PLG would impact their problem lists very little. It is also interesting to note that some students predicted that the PLG would help them a great deal, while others predicted that it would be detrimental to their lists. An interesting question to answer would be, how well does student estimation of the effect of the PLG on the quality of their problem lists actually predict the quality of their problem lists.

Clarity and Feasibility

At the time the field trial period began, it was known that the PLG would run reliably on Internet Explorer 4.0 for the PC, and was believed that the PLG would run well on Netscape Communicator 4.X for the PC as well. Ironically, the only major platform and browser combination not supported by the Vet School's computer lab was Internet Explorer 4.0 for the PC. During the field trial period it was also discovered that, while the PLG appeared to run correctly on Netscape 4.X, Netscape would cause major system crashes when used with longer cases. Furthermore, many students were unable to determine which browser they were using (or, if they could tell which browser, they had a hard time ascertaining the version). This set of circumstances frustrated many students a great deal before they even had the opportunity to start the PLG. It seems naïve to assume that this set of circumstances did not affect the questionnaire results.

Second, because the programmer had not finished the PLG completely at the time we ran the field trial, there were still some bugs in the program. These bugs and workarounds were well-documented in a help screen accessible from the PLG main page, but many either failed to read the help page, or did not understand it, and consequently were very frustrated by known bugs. (The lowest item on the entire questionnaire is the one dealing with technical difficulty).

Third, many of the criticisms the students had of the interface were criticisms of features that made their learning task a little more difficult by design. For example, students are forced to spell data abnormalities correctly with the PLG, which was annoying to some. I believe that, overall, the feedback concerning clarity and feasibility might best be summarized as follows:

1. Some of the frustration experienced by students and reflected in the questionnaire can be considered "productive" or, at least "planned." The fact that students are required to identify all data abnormalities before creating a problem list, or to finish a problem list before submitting a diagnosis are examples of this.
2. Fixing the bugs and making some relatively minor usability enhancements should improve ratings.
3. The PLG was tested in the field trial for Groups A and B (the groups whose data was used to indicate feasibility and impact) as a "walk up and use" application. This means that no formal effort was made to show students how the software worked at any point. This led to some frustration, particularly where bugs were concerned, which was made worse by the fact that students were not able and/or willing to find and use the help file. This same frustration was not present with the Group C learners, who received very little help, but for whom help was available immediately. Given this information, it seems wise to have the professor "demo" a case at the beginning of the semester so that students can see how it works and have questions answered.

Conclusion and Suggestions for Further Evaluation

In a keynote address at the 1999 Professors of Instructional Development and Technology (PIDT) meeting, Dr. Vic Bunderson (1999) of Brigham Young University proposed that most instructional software requires at least four cycles of implementation, evaluation and revision to achieve its full effectiveness. Therefore, he suggested that evaluations be planned to measure the effects of instructional tools through a number of implementation cycles. Given that perspective, it seems safe to say that the PLG has emerged, quite successfully, from the first cycle. However, there are several important questions that have yet to be answered, and changes that emerge from this cycle are sure to affect the results of the next (and first full) implementation cycle. We have shown that in the short term students who are accustomed to creating problem lists without the PLG produce problem lists that are just as accurate and complete using the PLG as with their more conventional methods. However, the PLG was designed to be a teaching and practice tool with long-term positive effects on student learning. The evaluation reported here was not able to measure the long term effects of using the PLG, nor was it able to compare long term PLG use to long term use of some other method of dealing with clinical laboratory data. Clearly, some of the comparisons made in this evaluation (such as the time comparisons) could logically be expected to produce results more favorable to the PLG if the students were as familiar with the PLG as they were with the other methods being used. Ideally, future summative evaluation efforts could involve robust experimental designs in which long term use of the PLG were compared to other treatments. However, there are many constraints that make this alternative impossible in the PLG's current context.

A good alternative, then, (and perhaps a better one) is to adopt Bunderson's far-sighted perspective. There are many ways of determining the effectiveness of the PLG

without comparing it directly to other instructional/practice methods. Student problem lists can always be compared to expert lists, and general discrepancies between them can be measured and used to guide improvements. Student opinions and suggestions will continue to be a valuable source of information for improving the PLG as well. Through such methods, the PLG, which is already a demonstrably effective practice tool, and which also addresses the administrative purpose that led to its creation, can aid many future learners in their quests to become effective medical professionals.

WORKS CITED

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues [published erratum appears in *Acad Med* 1993 Aug;68(8):615] [see comments]. *Acad Med*, 68(1), 52-81.

Amos, E., & White, M. J. (1998). Teaching tools: Problem-based learning. *Nurse Educator*, 23(2), 11-14.

Amsel, A., & Rashotte, M. E. (1984). Mechanisms of adaptive behavior: Clark L. Hull's theoretical papers, with commentary. New York: Columbia University Press.

Bednar, A. K., Cunningham, D., Duffy, T. M., & Perry, J. D. (1995). Theory into practice: How do we link? In G. J. Anglin (Ed.), Instructional technology: past, present, and future (2 ed., pp. 100-112). Englewood, Colorado: Libraries Unlimited, Inc.

Bergus, G. R., Chapman, G. B., Gjerde, C., & Elstein, A. S. (1995). Clinical reasoning about new symptoms despite preexisting disease: Sources of error and order effects. *Family Medicine*, 27(5), 314-320.

Berkson, L. (1993). Problem-based learning: Have the expectations been met? *Acad Med*, 68(10 Suppl), S79-88.

Bligh, J., & Wilkinson, P. (1997). Report of a workshop on problem-based learning and its implications for medical education in the UK. *Postgraduate Medical Journal*, 73(861), 449-459.

Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Acad Med*, 69(11), 883-885.

Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Acad Med*, 66(9 Suppl), S70-72.

Bridgham, R., Solomon, D., & Haf, J. (1991). The effect of curriculum era on NBME part 1 outcomes in a problem-based versus a traditional curriculum track. *Acad Med*, 66(9 Suppl), S82-84.

Brockett, R. G., & Hiemstra, R. (1991). Self-direction in adult learning: Perspectives on theory, research, and practice. New York: Routledge.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.

Bruner, J. (1990). Acts of meaning. Cambridge, Massachusetts: Harvard University Press.

Bunderson, C. V. (1999) Foundations of Design Experiments, Paper presented at the annual conference of Professors of Instructional Design and Technology (PIDT), Estes Park, CO. May, 21, 1999.

Burton, J. K., & Magliaro, S. (1988). Computer programming in generalized problem-solving skills: In search of direction. In W. M. Reed & J. K. Burton (Eds.), Educational computing and problem solving. New York: The Haworth Press.

Burton, J. K., Moore, D. M., & Magliaro, S. G. (1996). Behaviorism and instructional technology. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (pp. 46-73). New York: Simon & Schuster Macmillan.

Claplow, J. A. H., Donaldson, J. F., Kardash, C., & Hosokawa, M. (1997). Learning in a problem-based medical curriculum: Students' conceptions. Medical Education, 31(6), 440-447.

Cholowski, K. M., & Chan, L. K. S. (1992). Diagnostic reasoning among second-year nursing students. Journal of Advanced Nursing, 17(10), 1171-1181.

Christensen, C., Heckerling, P. S., Mackesy, M. E., Bernstein, L. M., & Elstein, A. S. (1991). Framing bias among expert and novice physicians. Acad Med, 66(9 Suppl), S76-78.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. Educational Psychology Review, 3(3), 149-210.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453-494). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cronbach, L. J., & Snow, R. E. (1981). Aptitudes and instructional methods: A handbook for research on interactions. (2 ed.). New York: Irvington Publishers Inc.

Cross, K. P. (1981). Adults as learners. (1st ed.). San Francisco: Jossey-Bass.

CTGV. (1990). Anchored instruction and its relationship to situated cognition. Educational Researcher, 19(6), 2-10.

CTGV. (1993). Anchored instruction and situated cognition revisited. Educational Technology, 33(3), 52-70.

Dick, W., & Carey, L. (1996). The systematic design of instruction. (4th ed.). New York, NY: HarperCollins College Publishers.

Dolmans, D. H., Gijsselaers, W. H., Schmidt, H. G., & Van Der Meer, S. B. (1993). Problem effectiveness in a course using problem-based learning. Academic Medicine, 68(3), 207-213.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (pp. 112-142). New York: Simon and Schuster Macmillan.

Estes, W. K. (1950). Toward a statistical theory of learning. Psychological Review, 57, 95-107.

Ferguson, D. B., & Rutishauser, S. C. B. (1997). A problem-based preclinical course for dental students. British Dental Journal, 182(10), 387-392.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1992). Principles of instructional design. (4th ed.). Fort Worth: Harcourt Brace Jovanovich College Publishers.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denszin & Y. S. Lincoln (Eds.), Handbook of qualitative research. Thousand Oaks, CA: Sage.

Gustafson, K. L., & Powell, G. C. (1991). Survey of instructional development models with an annotated ERIC bibliography (Information Analyses- ERIC Clearinghouse Products (071) -- Reference Materials -- Bibliographies (131) ED 335 027). Syracuse, NY: Syracuse University.

Hix, D., & Hartson, H. R. (1993). Developing user interfaces: Ensuring usability through product and process. New York: John Wiley & Sons, Inc.

Hix, D., & Hartson, H. R. (1997). Usability engineering course packet: CS 5714: Fall 1997. Blacksburg, VA: Virginia Tech.

Itani, K. M. F., Miller, C. C., Church, H. M., & McCollum, C. H. (1997). Impact of a problem-based learning conference on surgery residents' in training exam (ABSITE) scores. Journal of Surgical Research, 70(1), 66-68.

Kaufman, D. M., & Mann, K. V. (1996). Comparing students' attitudes in problem-based and conventional curricula. Academic Medicine, 71(10), 1096-1099.

Kearsley, G. (1998). Explorations into learning and instruction: The theory into practice database, [Web Site]. Available: <http://www.gwu.edu/~tip/index.html> [1998, June - September].

Kennedy, S. W., & Wilkerson, L. (1993). Topics for discussion reported by students using clinical cases in problem-based tutorials during a basic science course. Acad Med, 68(10 Suppl), S31-33.

Kindelan, J., Roberts-Harry, D., & Luther, F. (1997). A study to determine if changes to an undergraduate orthodontic course improved course quality and student satisfaction. European Journal of Dental Education, 1(3), 138-142.

Knowles, M. (1984). Andragogy in action. San Francisco: Jossey-Bass.

Lancaster, C. J., Bradley, E., Smith, I. K., Chessman, A., Stroup-Benham, C. A., & Camp, M. G. (1997). The effect of PBL on students' perceptions of learning environment. Academic Medicine, 72(10), S10-S12.

Lindsay, P. H., & Norman, D. A. (1977). Human information processing: An introduction to psychology. (2 ed.). Orlando, Florida: Academic Press.

Lyman, G. H., & Balducci, L. (1994). The effect of changing disease risk on clinical reasoning. Journal of General Internal Medicine, 9(9), 488-495.

Mandin, H., Jones, A., Woloschuk, W., & Harasym, P. (1997). Helping students learn to think like experts when solving clinical problems. Acad Med, 72(3), 173-179.

Merriam-Webster. (1998). WWWebster dictionary [on-line]. Merriam-Webster. Available: <http://www.m-w.com/mw/netdict.htm> [1998, April 1998].

Merrill, M. D., & Group, I. R. (1996). Instructional transaction theory: Instructional design based on knowledge objects. Educational Technology, 36(3), 30-37.

Merrill, M. D., & Tennyson, R. D. (1977). Teaching concepts: An instructional design guide. Englewood Cliffs, New Jersey: Educational Technology Publications.

- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. The Psychological Review, 63(2), 81-97.
- Morton, A. (1977). A guide through the theory of knowledge. Encino, California: Dickenson Publishing Company, inc.
- Norman, D. A. (1982). Learning and memory. San Francisco: W.H. Freeman.
- Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: A review of the evidence. Acad Med, 67(9), 557-565.
- Patel, V. L., Groen, G. J., & Norman, G. R. (1991). Effects of conventional and problem-based medical curricula on problem solving. Acad Med, 66(7), 380-389.
- Rand, J. S., & Baglioni, A. J. J. (1997). Subject-based problem-based learning in the veterinary science course at the University of Queensland. Aust Vet J, 75(2), 120-125.
- Rivarola, V. A., Bergesse, J. R., Garcia, M. B., & Fernandez, A. C. (1997). A problem-based learning approach to a biological chemistry laboratory class for students of veterinary medicine. Biochemical Education, 25(1), 22-23.
- Rosing, J. (1997). Teaching biochemistry at a medical faculty with a problem-based learning system. Biochemical Education, 25(2), 71-74.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: Theory and implication [published erratum appears in Acad Med 1992 Apr;67(4):287]. Acad Med, 65(10), 611-621.
- Schwartz, R. W., Donnelly, M. B., Sloan, D. A., & Strodel, W. E. (1997). Residents' evaluation of a problem-based learning curriculum in a general surgery residency program. American Journal of Surgery, 173(4), 338-341.
- Shneiderman, B. (1992). Designing the user interface: Strategies for effective human computer interaction. Reading, MA: Addison-Wesley.
- Skinner, B. F. (1953). Science and human behavior. New York,: Macmillan.
- Skinner, B. F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24(2), 86-97.
- Skinner, B.F. (1974). About behaviorism. New York: Knopf.
- Speelman, C. (1998). Implicit Expertise: Do we expect too much from our experts? In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, M. Anderson, & C. MacLeod (Eds.), Implicit and explicit mental processes (pp. 135-147). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Stevens, R. H. (1991). Search path mapping: A versatile approach for visualizing problem- solving behavior. Acad Med, 66(9 Suppl), S73-75.
- Technology, E. (1993). Landamatics ten years later: An interview with Lev N. Landa. Educational Technology, 33(6), 7-18.
- Tessmer, M. (1993). Planning and conducting formative evaluations: Improving the quality of education and training. London: Kogan Page.
- Tessmer, M., & Richey, R. C. (1997). The role of context in learning and instructional design. Educational Technology Research and Development, 45(2), 85-115.

Thorndike, E. L. (1971). The fundamentals of learning. New York,: AMS Press.

Van Treuren, K. W., & Havener, A. G. (1997). Teaching introductory engineering: A problem based learning experience. Biomed Sci Instrum, 34, 1-6.

Vasconez, H. C., Donnelly, M. B., Mayo, P., & Schwartz, R. W. (1993). Student perceptions of the effectiveness of a problem-based surgery curriculum. Acad Med, 68(10 Suppl), S28-30.

Walton, J. N., Clark, D. C., & Glick, N. (1997). An outcomes assessment of a hybrid-PBL course in treatment planning. Journal of Dental Education, 61(4), 361-367.

Wertsch, J., & Rogoff, B. (1984). Editor's Notes. In J. Wertsch & B. Rogoff (Eds.), Children's learning in the zone of proximal development (pp. 1-6). San Francisco: Jossey-Bass.

Wetherell, J., Mullins, G., Winning, T., & Townsend, G. (1996). First-year response to a new problem-based curriculum in dentistry. Australian Dental Journal, 41(5), 351-354.

Wilkerson, L., Hafler, J. P., & Liu, P. (1991). A case study of student-directed discussion in four problem-based tutorial groups. Acad Med, 66(9 Suppl), S79-81.

Winn, W., & Snyder, D. (1996). Cognitive perspectives in psychology. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (pp. 112-142). New York: Simon & Schuster Macmillan.

APPENDIX A -- INTERFACE DESIGN

Introductory Summary

The PLG will provide a tool for allowing students to classify lab data. It is hoped that students using this tool will develop the ability to classify data using the expert scheme upon which the tool is based. The expert scheme referred to, which will be described in some detail later, basically entails organizing data hierarchically in a way that unambiguously identifies causal relationships between bits of information.

Description of How the System Works

Overall Description

This tool will be accessible via a web browser. It must be able to interface with a database (much of the case information is already in a FileMaker Pro database.) Because this will be used as a grading tool (students are required to complete each case for credit), and because some of the images are not appropriate for a general set of viewers, access to the database will be password protected.

The menus, as explained in Figure A.1, will contain the following: File Menu: “Quit”; Help Menu: “Index”, and “About”. In Figure A.2, an introductory window is shown superimposed over a Macintosh background. Throughout this program, the menu will remain fixed, while the windows will be able to be dragged, minimized, and closed

The introductory screen (Figure A.1) is the login screen. individually. In this window, very few options are available, because the user has not yet logged in. Each user will log-in by entering a unique name and password. Any field receiving a mouse click will be active (show a flashing mouse cursor). The user can also move the active cursor from the name field to the password field by pressing the “tab” or “return” keys. To submit the information in the fields, the user must click the “Logon” button. If the name and password entered by the user are recognized by the system, the “Open Case” Screen will appear as shown in Figure A.2. If the user enters a user name which is not recognized by the system, a dialogue box with an “okay” button will appear reading “The user name entered does not appear in the database.” If a valid user name is entered, but the password is not correct, a dialogue box with an “okay” button will appear which reads “The password you entered doesn’t match your user name.” In either case, clicking the “okay” button will result in the dialogue box disappearing, and the student returning to the main logon screen.

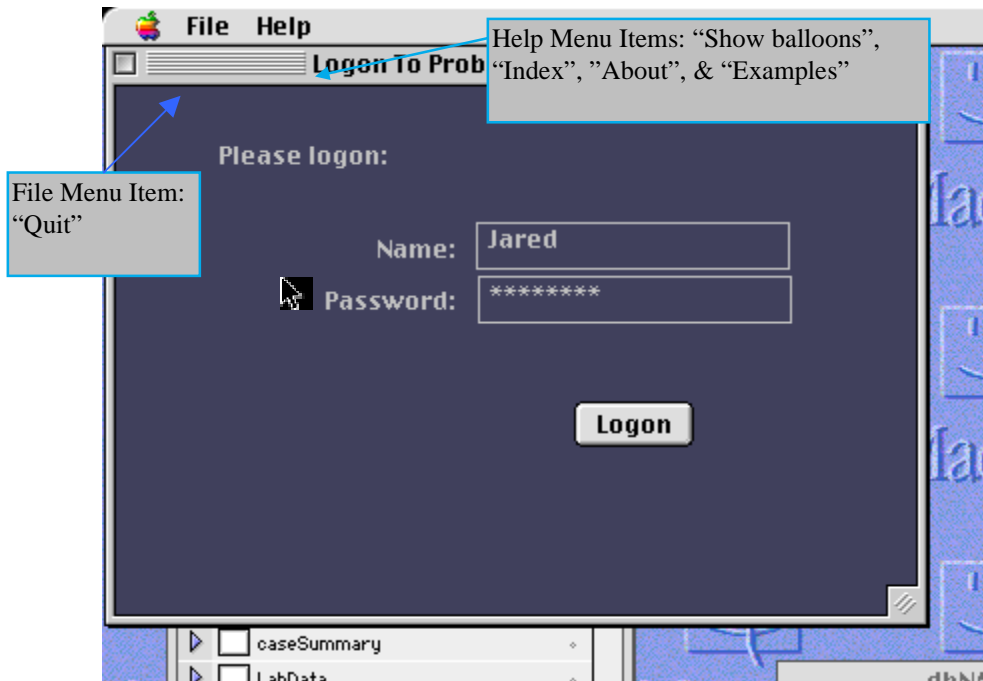


Figure A.1. Introductory Screen

Figure A.2, below, shows the initial screen once the user has successfully logged on. The initial screen is left purposefully simple. The toolbar and menu are redundant. Everything in the toolbar is also in the menu. The menu items contain the following:

Case Menu: Open, Close, Save, Find, Print, Evaluate, & Exit

Edit Menu: Undo Last Action, Copy, Paste, Select All, & Delete

View Menu: Case Summary, Lab Data, My Problem List, Expert Problem List, and My History (The Expert Problem list is grayed out on the toolbar and in the menu until the user problem list is submitted.)

Help Menu: Index, About, & Show Help Labels

The toolbar contains the following items (see Figure A.2 below): Open, Print, Save, Help, Search, View Lab Data, View Expert Problem List, View Learner Problem List, & View Case Summary.

The toolbar will appear at the top of the screen and will always be visible. The icons on the toolbar will be representative of the concepts being shown. The icons which might be most problematic to represent are the “expert problem list”, “my problem list”, “lab data” and “case summary” icons. In this document, the expert problem list is represented by an “expert” lecturing at a chalkboard, the learner’s problem list (my problem list) is represented by a pencil writing a list, the lab data is represented by a microscope, and the case summary is represented by a small picture of an animal representative of whatever case is being viewed. This image before a case is opened will be a generic image of an animal (such as the cat shown.) If a user tries to do anything other than open a case, or

view their history from this screen, a dialogue box will appear stating “You have not yet opened a case,” with two buttons reading “open case” and “cancel”. The “Open case” button will execute the open command, as described in the next paragraph, and the “cancel” button will make the dialogue box disappear.

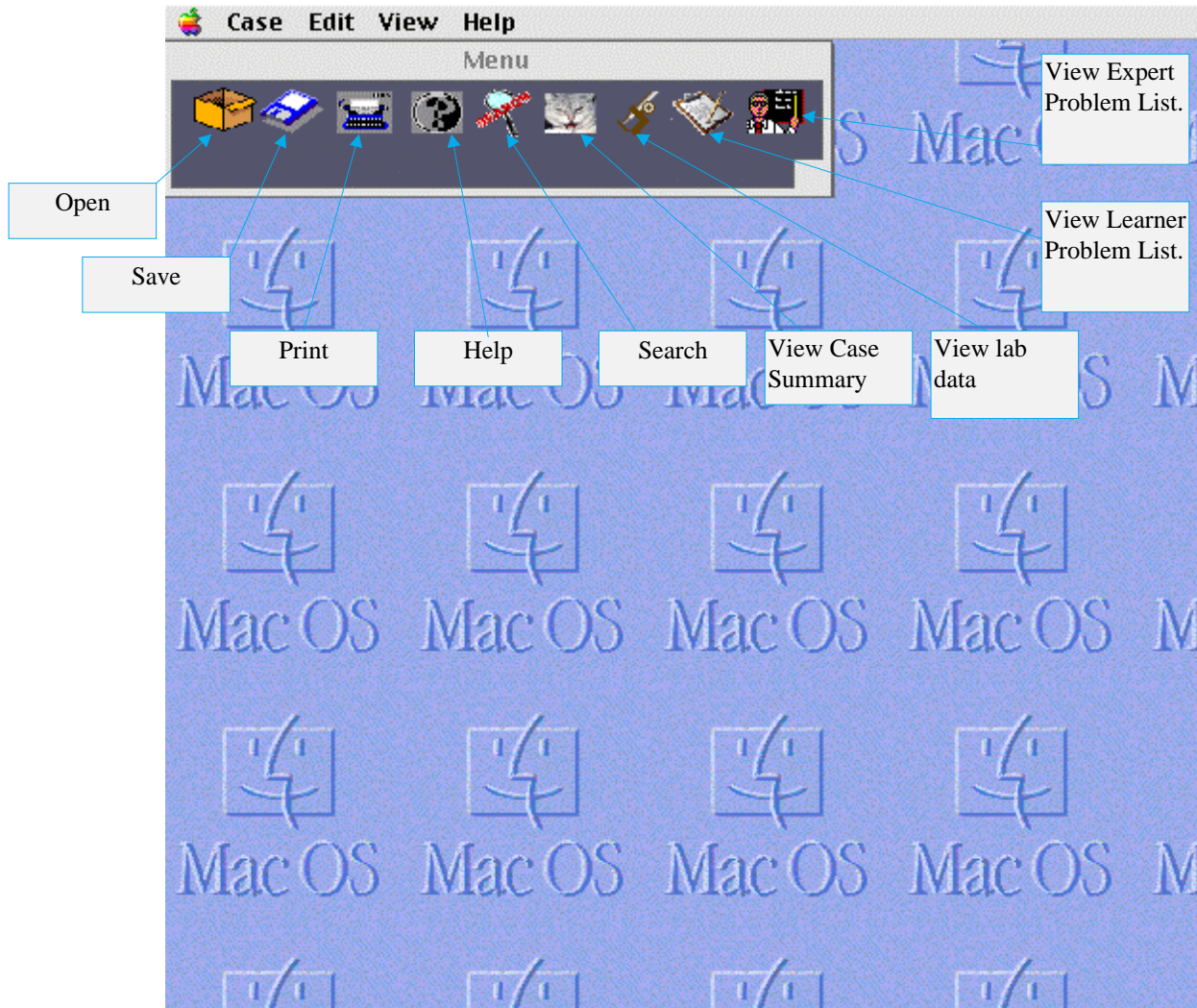


Figure A.2. Initial Screen After Login

The “open” command will cause a drop-down menu with openable cases to appear, as seen in Figure A.3 below. Once the learner has chosen a case, all items under the view menu will refer to that case, except for “history” which will refer to all cases that have been completed by the user. Cases will be listed in a pull-down menu.

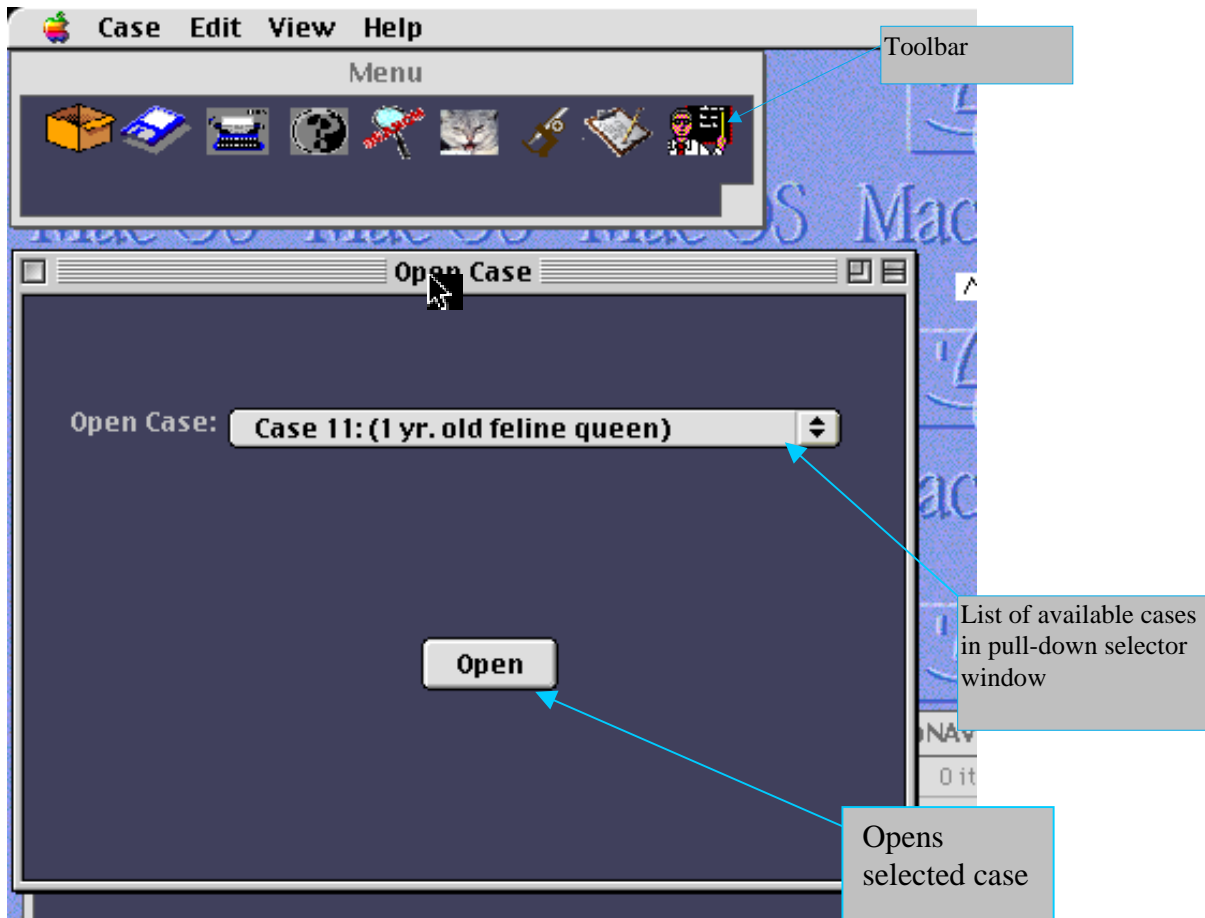


Figure A.3. "Open" Screen

The professor can choose to make certain cases available or unavailable. Unavailable cases will be grayed out. Once a case has been clicked to open, the case summary screen, as seen in Figure A.4 below, will appear for the selected case.

At this point, all menu and toolbar options will be available except View: My Problem List, Case: Evaluate and View: Expert Problem List. View: My Problem List becomes available when all lab data abnormalities have been classified and named in the "lab data" window. View: Expert Problem List and Case: Evaluate become available when "My Problem List" has been submitted to the database by submitting a diagnosis.

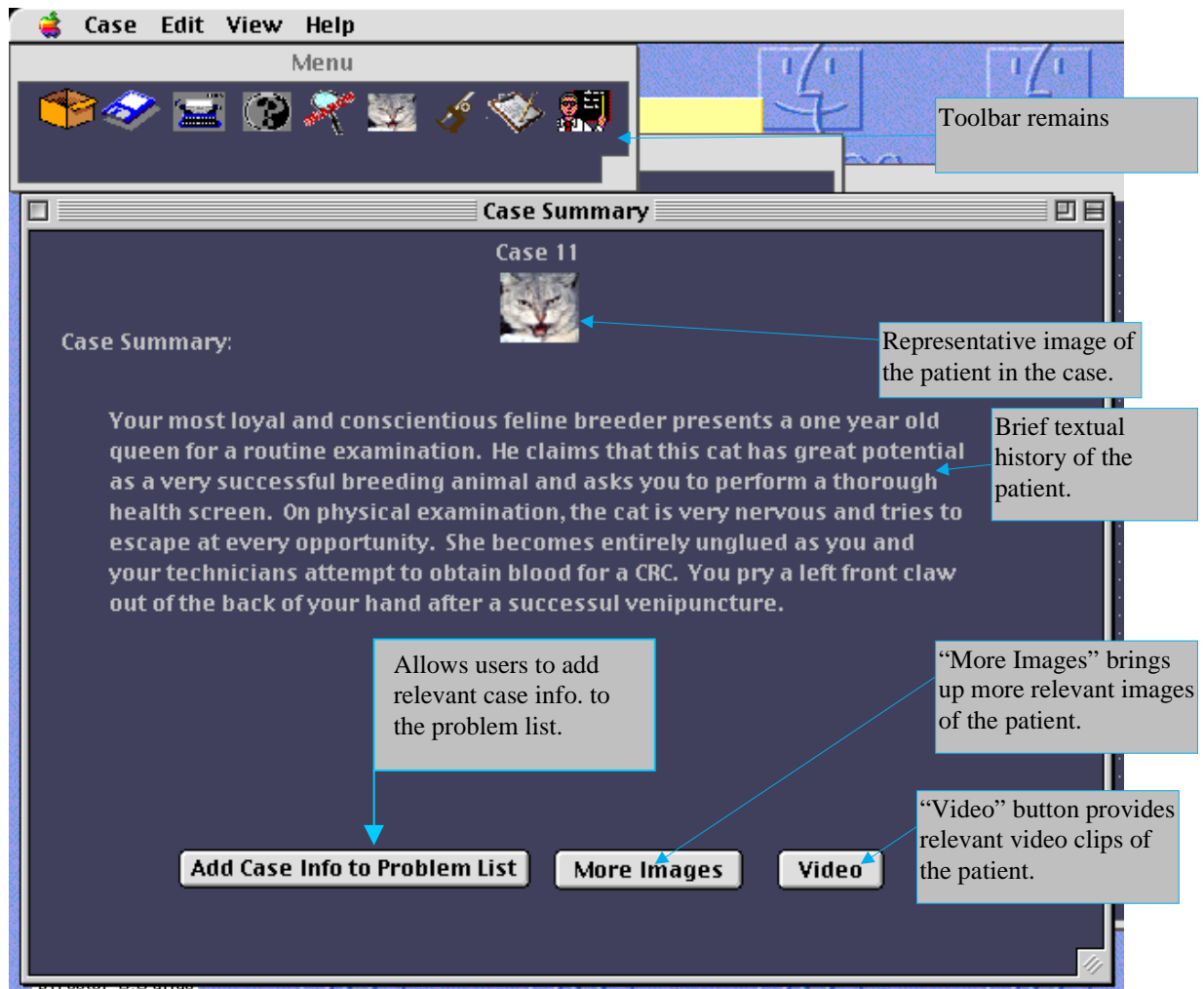


Figure A.4. Case Summary

Returning to the case summary screen (Figure A.4, above), it will sometimes be beneficial to view additional videos and/or images of the patient. In such cases the “More Images” and “Video” buttons (as seen at the bottom of Figure A.4) will be active. If there are no videos or images available, the respective button(s) will be grayed out. Images and videos, when selected, will appear in separate, closeable/minimizable windows. The videos will appear in a typical quicktime-like format that will allow users to play, pause, jump forward, and jump back in the video.

The “Add Info to Problem List” feature is designed to allow students to add relevant information from the case summary that would not be seen in an analysis of the lab data, but that would be useful in creating the problem list. Clicking the “Add Case Info to Problem List button” results in the dialogue box shown in Figure A.5 below. Once the user has entered a name for the problem or phenomenon identified, and has written a description of it (if desired), clicking the “Add to my problem List” button will result in the problem appearing in the “Data Abnormalities I’ve found” column of “My Problem List.” The notes will be active, and could be seen at that point, if toggled, by way of the

notes function, described later. If the user decides not to record the problem found, the cancel button will cause the dialogue box to disappear.

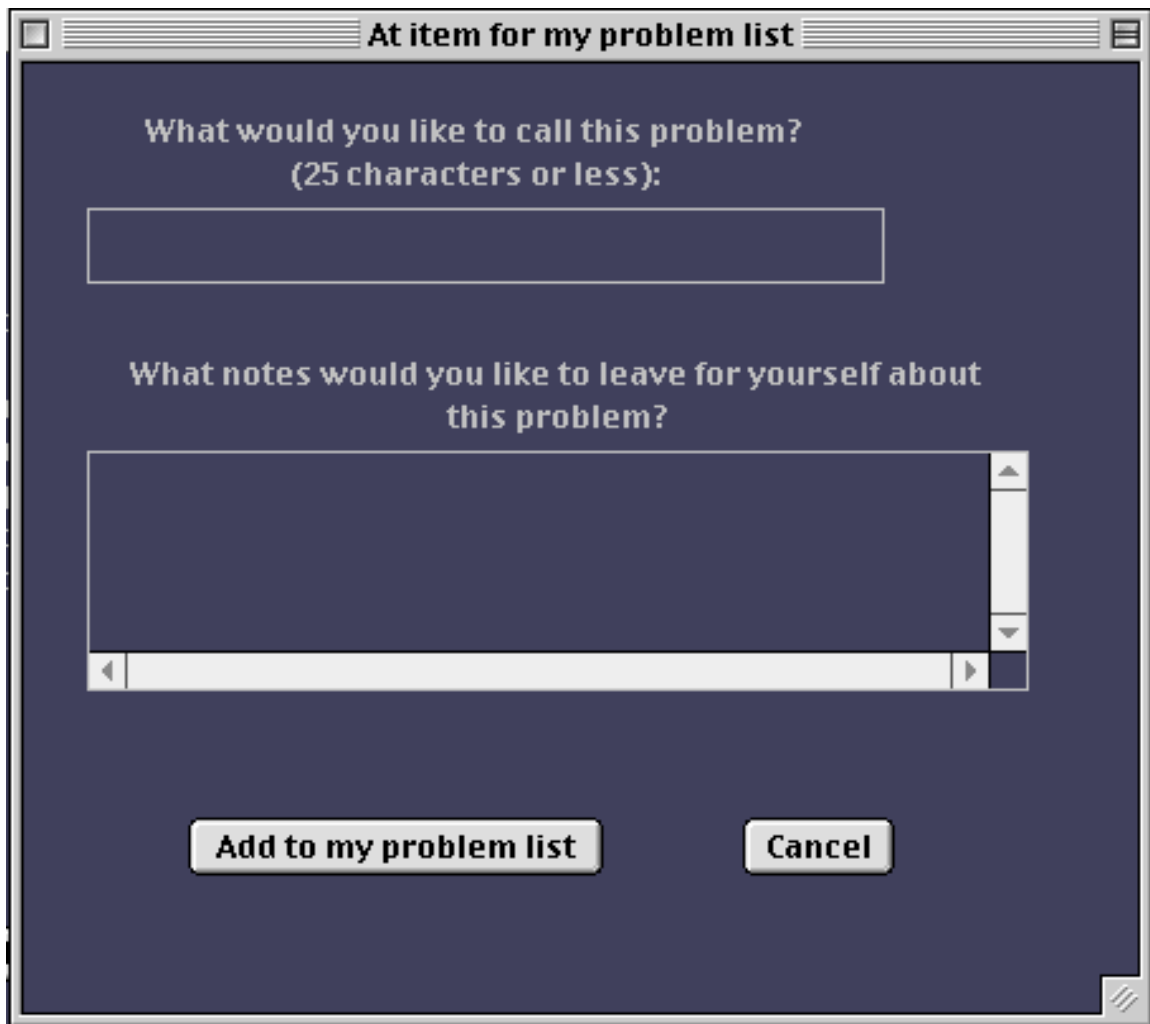
A screenshot of a dialog box titled "At item for my problem list". The dialog box has a dark blue background and a light blue border. At the top, the title "At item for my problem list" is displayed in a white font. Below the title, there are two questions in white text: "What would you like to call this problem?" followed by "(25 characters or less):". Below this question is a white rectangular input field. The second question is "What notes would you like to leave for yourself about this problem?". Below this question is a larger white rectangular area with a vertical scrollbar on the right and a horizontal scrollbar at the bottom, indicating a text area for notes. At the bottom of the dialog box, there are two buttons: "Add to my problem list" on the left and "Cancel" on the right. Both buttons have a light blue background and a dark blue border. In the bottom right corner of the dialog box, there is a small icon consisting of three diagonal lines.

Figure A.5. Adding an item to the problem list from the lab data screen

Data Classification and Synthesis

When the user chooses to view the lab data, a screen will appear as in Figure B.6 below. The purpose of the screen illustrated in Figure B.6 (below) is to allow students to classify the data, and name data abnormalities.

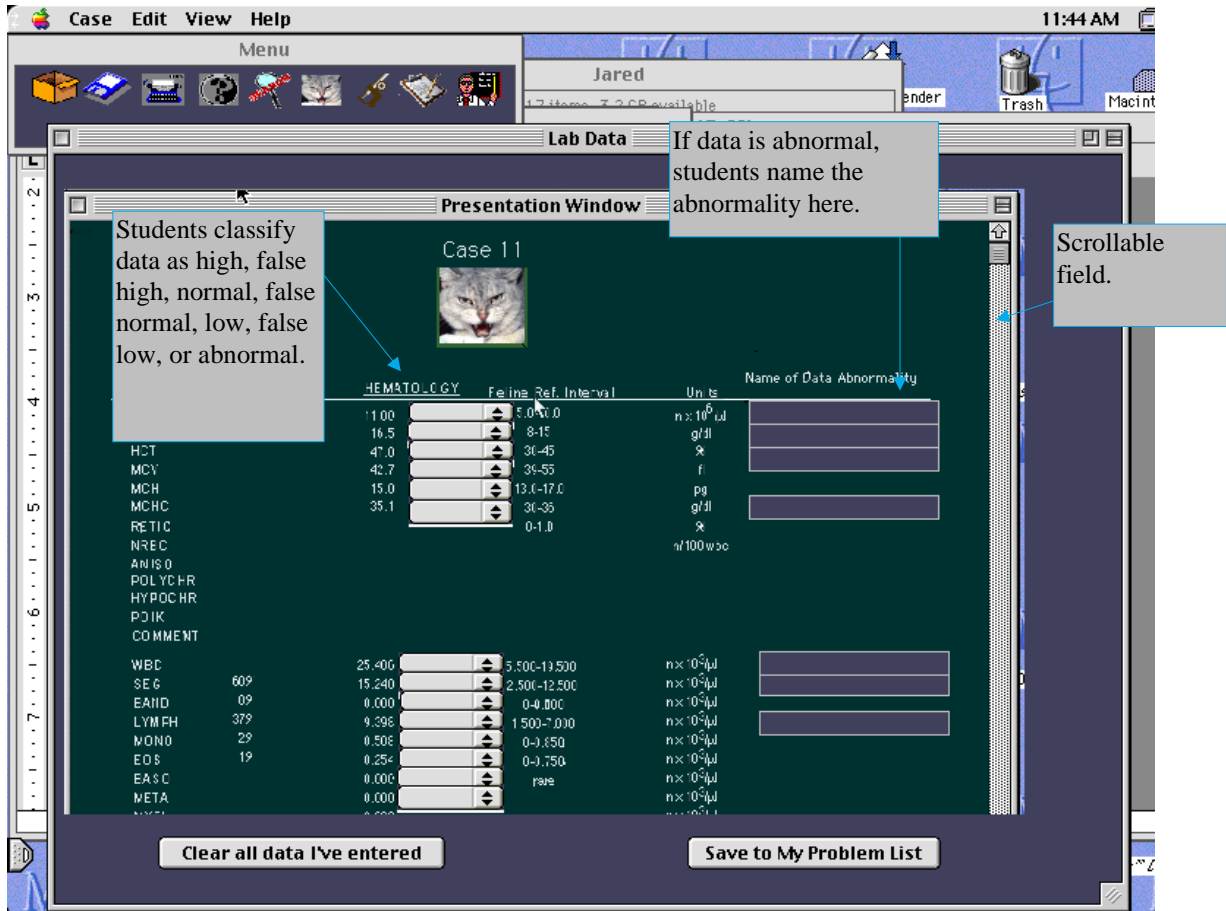


Figure A.6. Lab Data

Data is classified as normal, false normal, high, false high, low, false low, or abnormal. This will be done using the pull-down menu to the right of each data point containing the following items: nothing (normal), false \leftrightarrow (false normal), “ \uparrow ” (high), “false \uparrow ” (false high), “ \downarrow ” (low), “false \downarrow ” (false low), and “abnormal”. (Note: It is also common to see high, normal and low data represented by “.....|”, “...|...”, and “|.....” respectively. These are the symbols that will be used in the screen shots for this document). The pull-down menus will be blank (normal) by default.

When a data item is changed to indicate anything other than “normal”, a textbox will appear to the right with a flashing text cursor, where the student will type the name of the data abnormality identified. If a data abnormality name is entered which does not match the appropriate name in the database, a text box will appear saying: “The data abnormality you have entered doesn’t match the database.” The student will be able to press a “help” button, (which will produce a help dialogue with a description of the data in question, and a link to a related tutorial), a “cancel” button (which will clear the text entry box), or a “change the abnormality name” button, (which will cause the dialogue box to disappear, and the text box with the name in question to regain focus.) At the third unsuccessful attempt to name the data abnormality, the same dialogue box will appear, only this time, it will also include a button labeled “Tell me.” If the “Tell me”

button is clicked, the dialogue box will disappear, and another will appear stating “When [name of data] is [high or low] it is known as [name of data abnormality].”

The lab data screen, shown in Figure A.6, is purposefully laid out to reflect the printouts that veterinary practitioners are forced to work from in the field. Learning theory suggests that the more a practice environment approximates the environment in which the task will have to be performed, the more likely the knowledge and skills gained are to transfer to the work environment.

As can be seen in Figure A.6, the user has identified seven items as abnormal (all high), and could now name them in the text boxes which were created to the right. Names of various data abnormalities can be dragged from one text box to another, which will result in that name appearing in both text boxes. A “Clear all data I’ve entered” button appears at the bottom of this screen, which will make the data abnormalities screen appear as it did before the user altered it. This will also clear the “Data Abnormalities I’ve Found” box of “My Problem List”, if “My Problem List” has not yet been opened.

If a user attempts to open “My Problem List” before accounting for all data abnormalities in the “lab data” window, a dialogue box will appear saying “All data abnormalities must be identified before you can create your problem list” with an “okay” button.

Pressing the “okay” button will cause the dialogue box to disappear and the lab data window to reappear. Pressing the “cancel” button will cause the dialogue box to disappear, and the “Lab Data” window to regain the focus.

When students enter the name of a data abnormality in the field provided as shown in Figure A.6, the name of that abnormality appears in a “data abnormalities field” (alphabetically) as shown on the right hand side of Figure A.7, below.

Figure A.7 illustrates the window that will be used to construct a problem list. The two boxes (problem list and data abnormalities), while in the same window, can be narrowed or widened. Names of data abnormalities can be clicked and dragged from one box to another, and can be dragged from one place to another in either box.

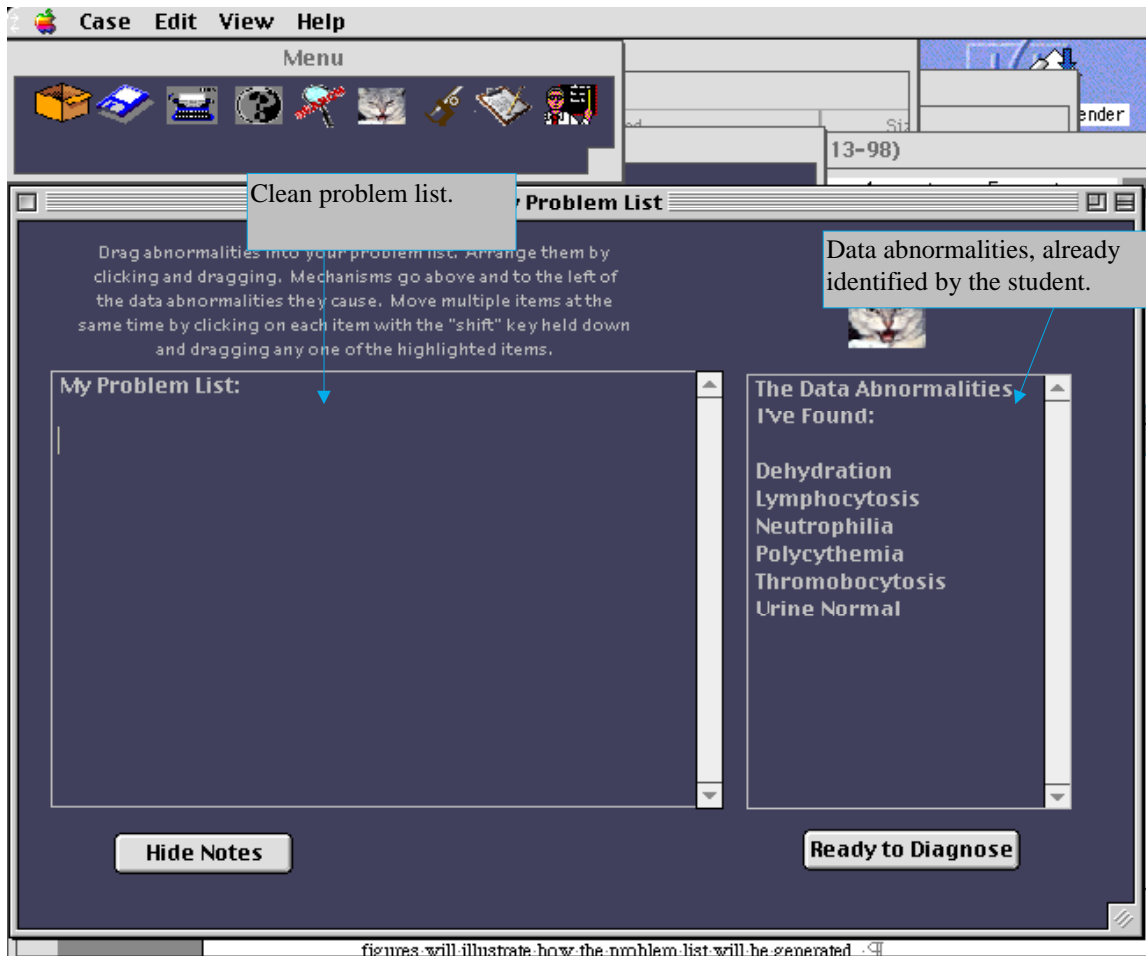


Figure A.7. The User's Problem List

The process of generating the problem list is at the heart of this product. The following figures will illustrate how the problem list will be generated. Figure A.8 (below) shows that several data abnormalities have been dragged from the "data abnormalities" box to the "My Problem List" box. Data abnormalities which have been dragged from the data abnormalities box continue to appear in the data abnormalities box, and will still be usable (i.e., any give data abnormality can appear in the problem list more than once), but they will appear with a check mark next to them in the Data Abnormalities box so that students will know they have been used at least once. We can also observe the "notes" function, which allows students to annotate each data abnormality, mechanism, or the diagnosis.

The notes feature functions much like the nesting structure in the system of a Macintosh. When the triangle is pointing to the right, the notes are concealed; when the triangle points down, the notes are revealed, and are editable. All data abnormalities and mechanisms will "carry" a "notes triangle", even if no notes have been entered. Toggling the triangle down for a mechanism or abnormality for which there are no notes, will reveal a one-line editable field for notes with a flashing cursor to indicate where to type. When students enter text, the text area will grow to five lines in length, and then will become scrollable, allowing students to enter as much text as they wish.

When a data abnormality is first dragged from the “Data Abnormalities I’ve Found” field to the “My Problem List Generation window, the notes field will appear open and active by default. The notes function will prove useful in printing and viewing the problem list -- students will be able to view as much or as little of their notes as they wish.

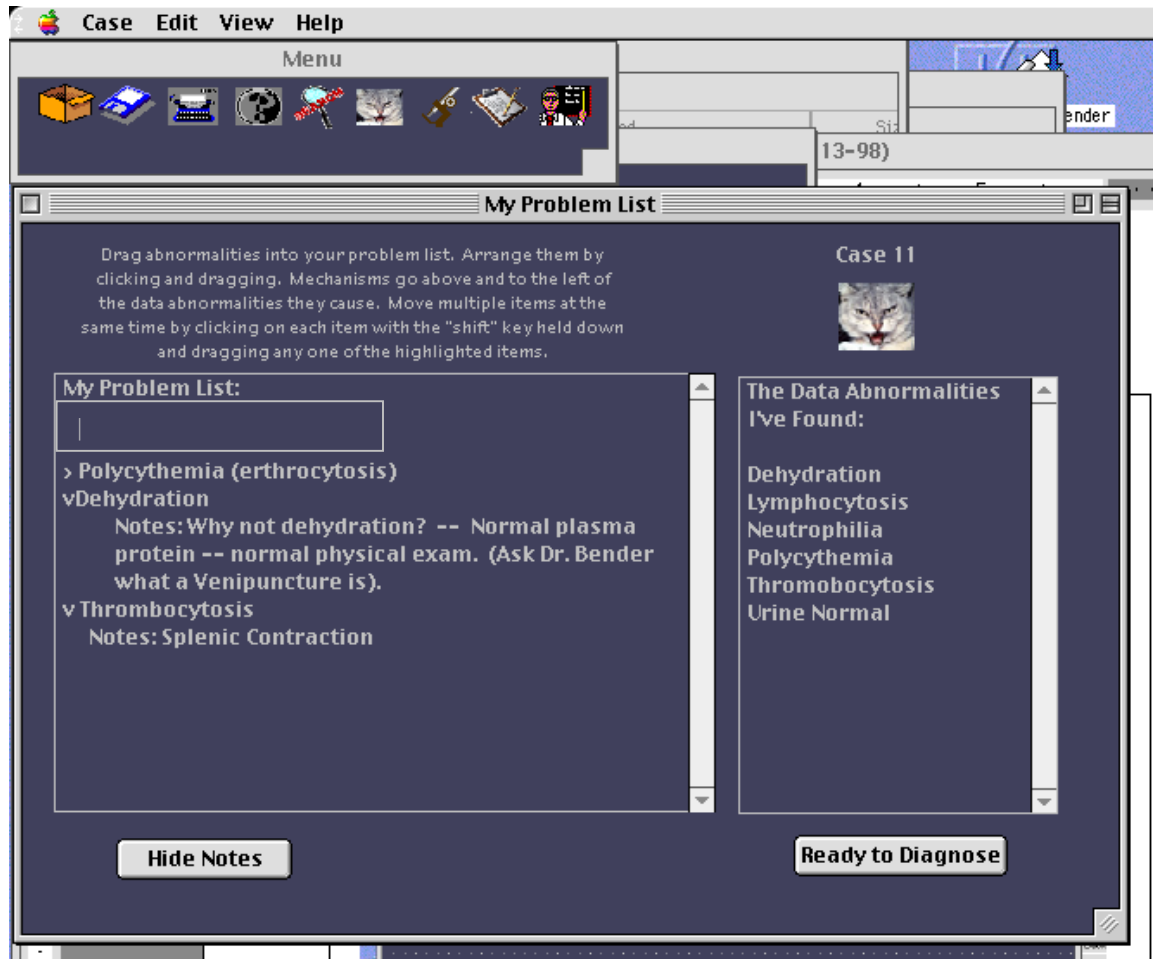


Figure A.8. Initial Problem List Generation

Data in the problem list will be organized hierarchically by level. A light grid will appear in the background of the problem list, indicating different hierarchical levels. It is assumed that data above and to the left of other data cause the data below and to the right to occur. When students first drag data abnormalities into the problem list, they can place them anywhere on the grid they choose. The student can choose to change the hierarchical level of abnormalities or mechanisms by tabbing them (tab to the right, shift/tab to the left), or by dragging them and dropping them directly to the right of any given grid line.

If one or more items of data are placed to the right of the furthest left position, and no mechanism appears above them, a text entry box will appear above and to the left of such data, where the student can enter the name of the mechanism which is

affecting the subordinate data. In Figure A.8, several pieces of data have been transferred to the problem list box, and have been annotated, and a text box has appeared above and to the left with a text cursor for entering the name of a mechanism. In Figure A.9 (below), the mechanism has been labeled “Splenic contraction.”

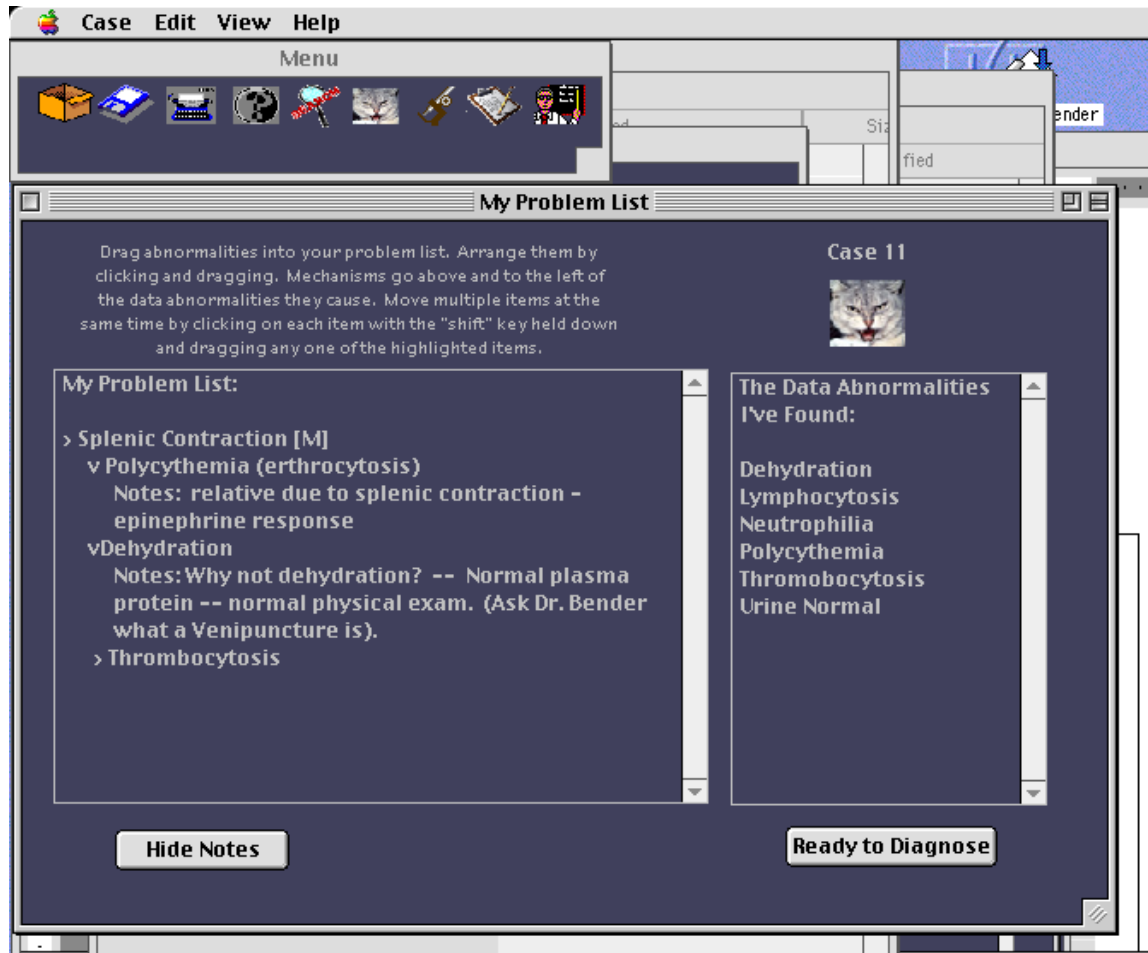


Figure A.9. My Problem List with Two Levels of Nesting

To the right of the words “Splenic Contraction”, in brackets, the student has toggled the label “M” (for mechanism). (To the right of each major term, a toggleable area will allow students to classify terms as mechanisms, data abnormalities, or “other”, using a stylized “M”, “DA”, and “Other”)

Note that this screen is very flexible. All items in the my problem list box can be dragged and dropped, above or below, or on top of any other item in the box. When an item is dragged on top of (or to the same cell as) another item, the two items are placed together, as one word (and are then, of course, editable). Items can also be dragged from the my problem list box back to the data abnormalities box if students decide they aren’t ready to deal with them yet, or if they decide that certain abnormalities, are, in retrospect, normal. Notes are not individually draggable, but remain associated with the data abnormalities or mechanisms to which they relate.

At this point, and throughout the entire process, all windows can be open simultaneously except the “Expert Problem List.” This option is grayed out of the “View” menu until a problem list and diagnosis have been submitted. Therefore, students can refer back to the laboratory data, reevaluate their decisions, identify new data abnormalities, refer back to the case summary, etc., throughout the entire process. The process of organizing data is iterative so it is expected that students will refer to each of the windows several times during the process of data classification. During this process students can arrange and open windows to best suite their individual needs.

Evaluation

When students have finished organizing the data to their satisfaction, they can choose to submit a diagnosis, by pressing the “Ready to Diagnose” button. When this occurs, a confirmation box will appear that states “When you submit a diagnosis, your problem list will be recorded, and you will no longer be able to alter it.”, with “Okay” and “Cancel” buttons. (Requiring that students submit an unalterable problem list before seeing the expert solution will ensure that they work through the problem completely before submitting it, rather than simply copying the expert solution.) “Okay” will be the default button (activated by the return key). When the “Cancel” button is pressed, the confirmation screen disappears. When “Okay” is pressed, a text entry box will appear in the same confirmation screen labeled “Diagnosis:”. When the student enters text in this box, and presses “Okay”, a second confirmation screen will appear which will say “Do you want to record ‘[contents of text entry]’ as your diagnosis? If the student selects “Cancel”, the first confirmation screen will come to the front again. If the student selects “Okay”, the problem list will be recorded by the system, and a dialogue box will appear which states “Your problem list has been submitted.” An “Okay” button will appear which, when pressed, will make the dialogue box disappear. The student’s diagnosis will appear at the top of the problem list, labeled “Diagnosis.”

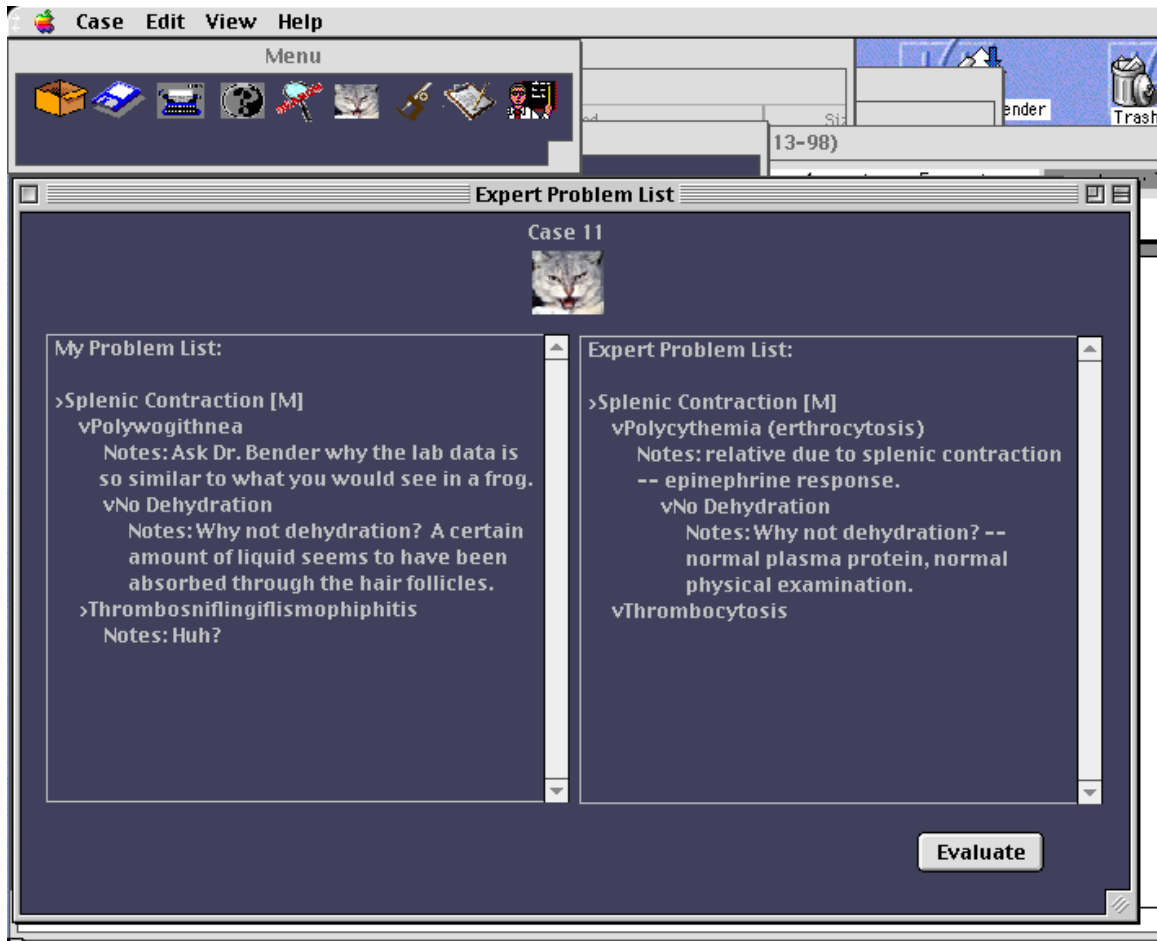


Figure A.10. “My Problem List” and “Expert Problem List” compared

At this point, the “Expert Problem List” item will also become active in the “View Menu.” The “Ready to Diagnose” button will now read “Evaluate” (as seen in Figure A.10, above) and will do the same thing as the “Evaluate” option in the “File” menu. Students can now compare their problem list with the expert problem list by positioning the screen for each problem list side by side, and looking for differences. The student will also be able to print and compare both lists, take them to classes to write notes on, etc.

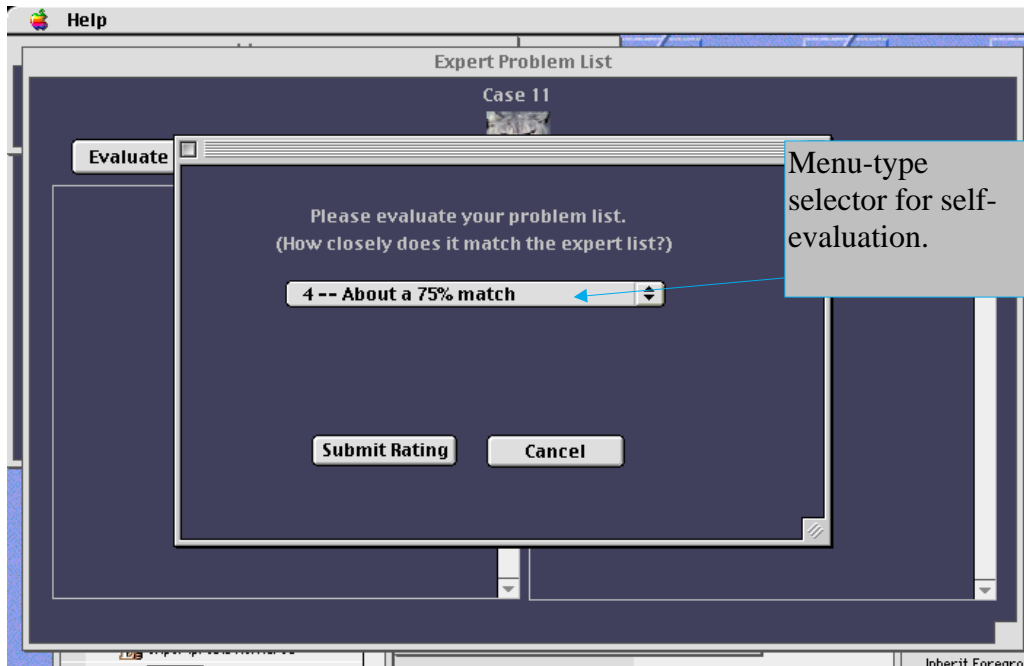


Figure A.11. Evaluation Box

Activating the “Evaluate” option, either through the button or the file menu, will cause a box to appear which reads “Please evaluate your problem list. How closely does it match the expert list?” (see Figure A.11, above). A pull-down menu will appear, allowing the student to choose one of the following categories (pull-down menu selector shown in Figure A.3 with option 4 selected): 1, Almost no match at all; 2, About a 25% match; 3, About a 50% match; 4, About a 75% match, & 5, A nearly 100% match. A menu-type selector will allow users to select which rating best characterizes the match between their problem list and the expert problem list. A “submit” button will allow users to submit their self-evaluations. A “cancel” button will make the evaluation box disappear.

Other Functions

Help

The “Help” function will provide support in two major areas: a. in the use of this program, and b. dealing with clinical data classification and synthesis. The “Help” Index will be broken up accordingly, containing two major areas of help which are then broken down further in a tree-like structure (see Figure A.11 below). As can be seen in Figure A.11, the two main help categories “Help with this program” and “Help with the content” are further broken down into sub-areas. “Help with this program” for example, contains the main sub-areas of “Case Summary”, “Expert Problem List”, “Help Function”, “Lab Data”, “My problem List”, “Opening a case”, “Printing”, and “Search function”. Those sub-areas are further broken down, in most cases. Double clicking a category will show its sub-categories, or, if it has no sub-categories, will open a text box describing how to use the function/object in question. The “Help” index section will provide the majority of the substance for all the help, including, where appropriate, pointers to relevant tutorials

(which are currently in the development process.) This will be accomplished by providing a “See related tutorial” button after the paragraphs describing how to use the function or object. The help is not fully developed at this time, and will require concerted effort on the part of the designer working with subject-matter-experts.

Print

There is an important instructional reason for allowing the users to print. First, because of past learning situations, learners often have learning strategies tied to how they manipulate their paper notes, etc. Second, many students will want a copy of their own and/or the expert solution when they go to class and hear the lectures, so that they can write down what they have learned from the lectures in a contextually significant way. This could help the learners relate what they learn with the classifier to what is happening in class.

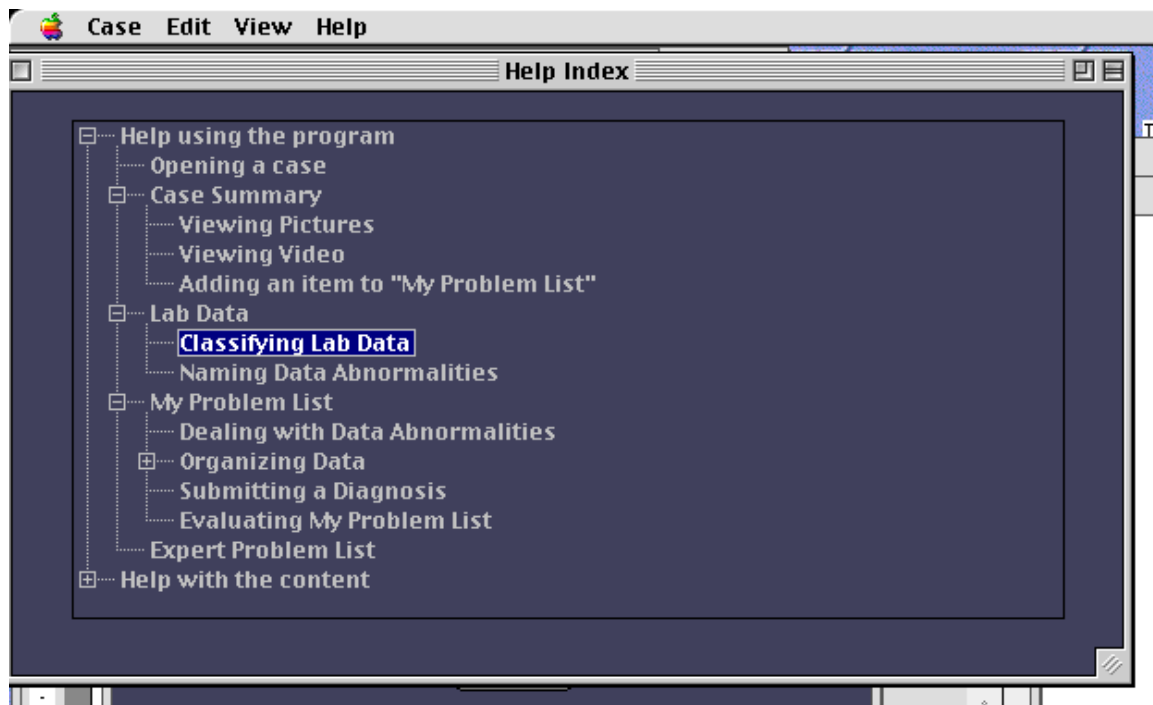


Figure A.12. Help Index

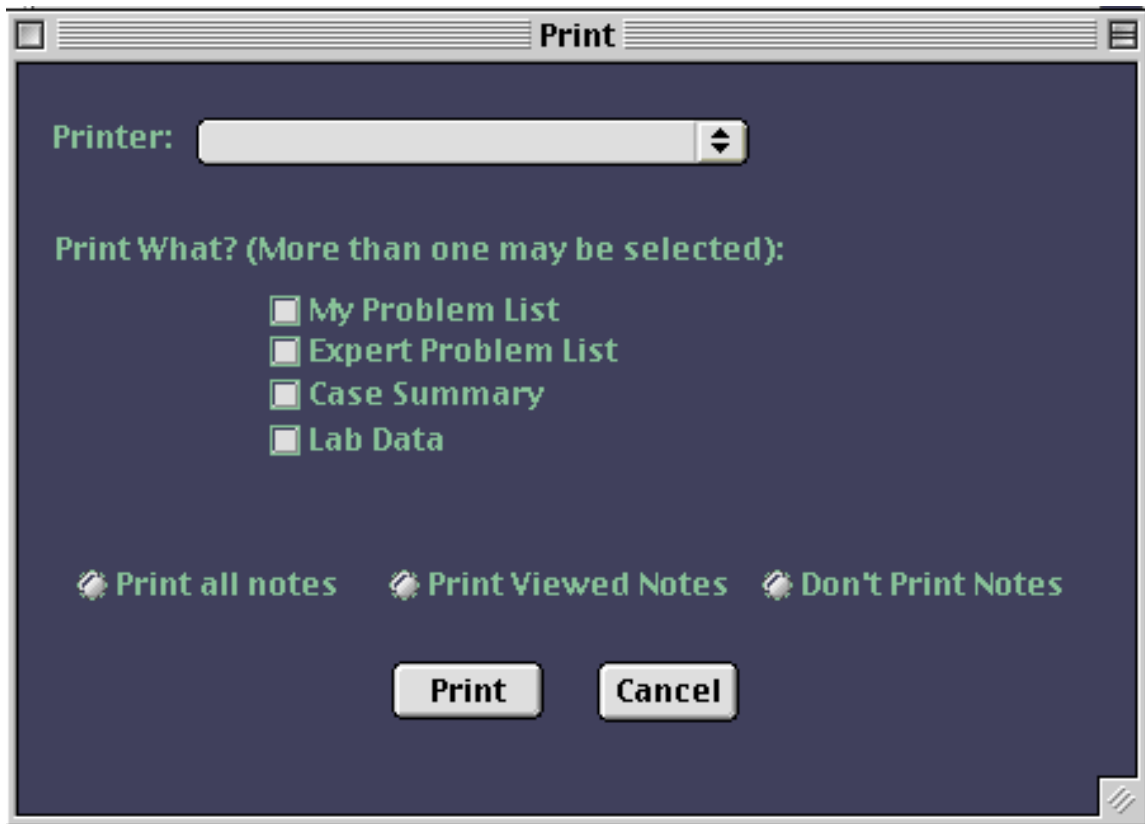


Figure A.13. Printer Interface

When the “print” option is selected (either from the pull-down menu, or from the menu bar), a printer dialogue box will appear as shown in Figure A.13, above.

As seen above, the printer interface will be relatively simple. Users can select all or none of the top four boxes. The bottom three radio buttons are mutually exclusive – selecting one deselects the other two. Choosing “print viewed notes” results in the selected lists being printed exactly as they appear in their respective windows, with the notes the user has left open being printed. Since Case Summary and Lab Data do not have notes, if they are the only options selected, the lower 3 options will be grayed out.

Search.

The search function will search for any given word or combination of characters, throughout the entire case, in any of the sub-screens selectable in the view menu.

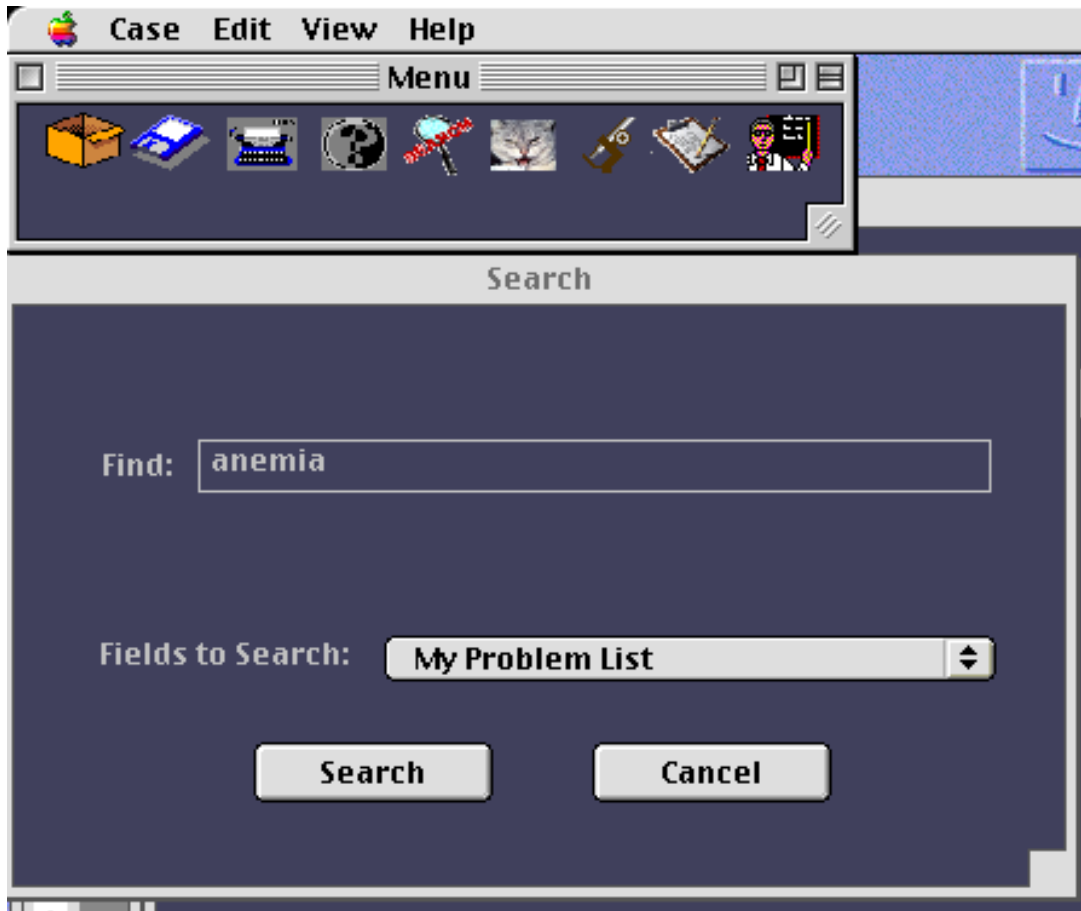


Figure A.14. Initial Search Screen

Students may want to find where they have placed a particular data abnormality or mechanism in a case, or they might want to search the expert problem list by name to identify where and how the experts have placed specific mechanisms or data abnormalities. The initial search screen will allow the user to identify which word will be searched for, and where the system should search, as seen in Figure A.14 above. Here the user has chosen to search for “anemia” in “my problem list.”

When the user clicks the “search” button shown in Figure A.14, a split “found” and “where” screen appear, as shown in Figure A.15, below. Any item appearing in the “found” box will have a corresponding item in the “where” box. When any item is double clicked, the window containing it will be opened, and the item will appear highlighted. In this case, for example, anemia has been found once in “My problem list” so double-clicking on anemia will open the “My Problem List” window where the word “anemia” appears, with the word “anemia” highlighted. If the word which has been searched for cannot be found, a dialogue box will appear with the words: “[The word searched for] was not found in [name of the location searched in],” with one button reading “Okay.”

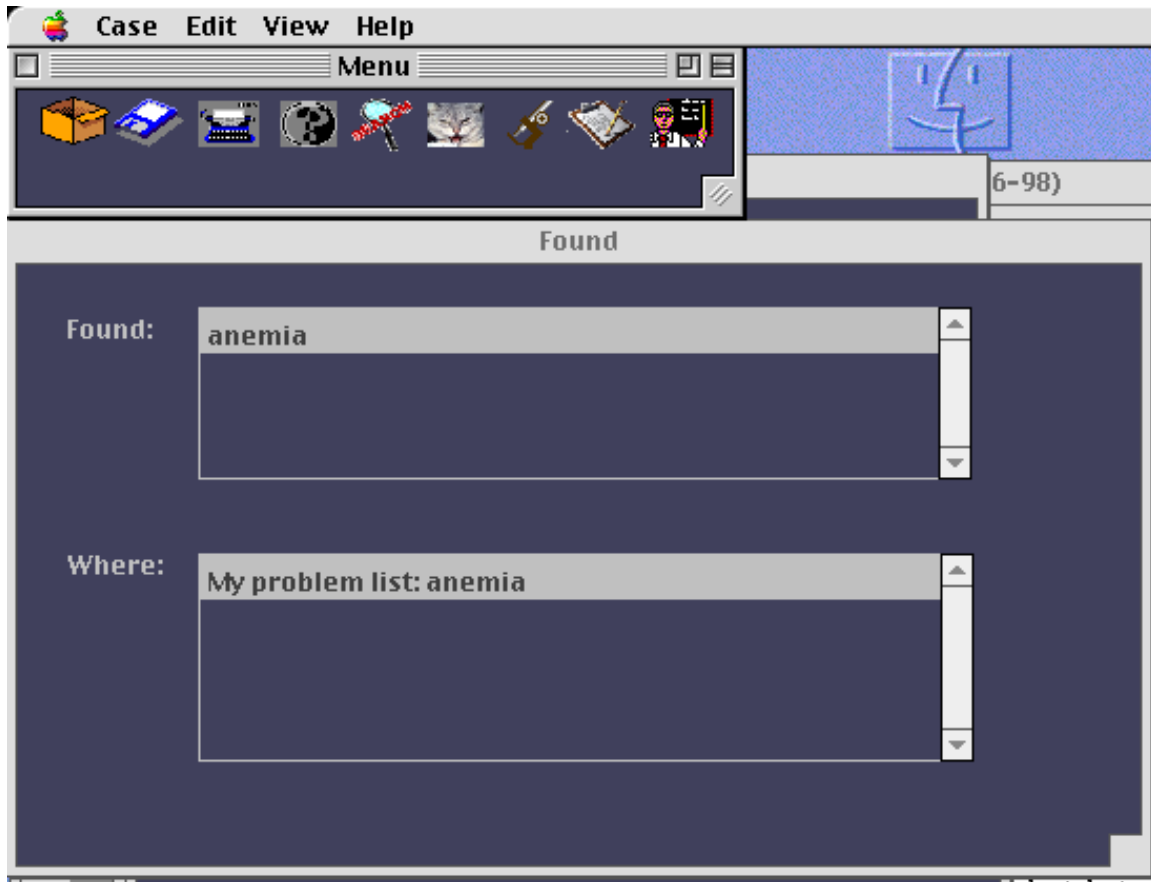


Figure A.15. Found Window

View my history

Choosing the “My History” option in the “View” menu will result in a screen which permits users to view how they rated themselves on each case. The scrollable screen will simply show each case number and description, the self-evaluation, and a “close” button at the bottom of the screen.

APPENDIX B - INSTRUCTIONAL DESIGN DOCUMENT

Instructional Design

Overview of the Problem List Generator and Instructional Problem

Veterinary students in VM 8414 “Clinical Pathology” are required for the first time in their classroom experience to synthesize the knowledge they have gained about anatomy, physiology, symptoms of medical problems, medical conditions, and characteristics of a variety of animals (species or breed-specific), in interpreting lab data and arriving at diagnoses. The students frequently find the process of data classification and synthesis difficult and frustrating. This problem can be seen as an instructional one because although the students have been taught the individual sets of knowledge involved in understanding the lab data, they lack the knowledge and skills to synthesize the data, classify it, and arrive at a diagnosis. Furthermore, they have forgotten a great deal of the prerequisite knowledge for this skill. A computer-based practice tool known as the Problem List Generator (PLG) will be design and implemented to help address this problem. The PLG is not meant to present the skills involved in problem classification. This will be achieved with related tutorials. Rather, the PLG will allow students to practice classifying data, and synthesizing it (which results in a problem list) and receive feedback.

The PLG will also address an administrative problem. The professor wants each student to complete and receive credit for completing a problem list for between 6-9 cases per week. There is currently no system in place to allow for this. The PLG will allow students to do the problems on their own, evaluate their own responses, and record the fact that they have completed the assignment.

Learner Analysis

Learners (general analysis):

- Age from 20 to 50+
- Approximately 70% Female/30% Male
- Veterinary students
- Highly motivated (This course is a degree requirement.)
- Highly competitive
- Traditionally successful in school situations
-

Contextual Analysis

Introduction to Context

For this contextual analysis I will rely on Tessmer and Richey's (Tessmer & Richey, 1997) conceptual framework for examining context. Their framework examines learner factors, immediate environment factors, and organizational environment factors in

each of three contexts: the orienting context, the instructional context, and the transfer context.

Learner factors are those factors which are individual to each learner. Immediate environment factors refer to the people and places immediately surrounding the students. Organizational factors are more removed from the student and include supervisory entities such as programs, departments, colleges, and administrative groups, as well as individuals within those entities.

The orienting context refers to the contextual factors that precede instruction. The instructional context refers to the contextual factors (time and place) directly associated with the delivery of the instruction. The transfer context refers to the environment where what was learned will be put in practice.

The following observations and implications are based on interviews with students, informal discussions with faculty members, and personal experience working in the Virginia Maryland Regional College of Veterinary Medicine (VMRCVM).

Orienting Context

Learner factors.

Naturally, learner factors vary as much as the learners themselves. In general, however, the learners at the VMRCVM have been selected for their ability to do well in formal educational settings (as evidenced by G.P.A.). In this particular domain of knowledge, learning success seems highly correlated with the ability to memorize verbal information and rules quickly.

Students at the VMRCVM are generally specifically "oriented" to the Clinical Pathology course by their predecessors. Most come into the class with the expectation that it will be very difficult and challenging, but that they will also learn a lot. *(Implication: If the students perceive the PLG as a tool that will save time, or somehow make their lives easier, their orientation will make them more likely to be excited about using it. On the other hand, if it makes an already difficult class more difficult, they are not likely to be motivated to use it.)*

Immediate environment factors.

Students' immediate orienting environment is their environment away from the learning environment. Tessmer and Richey describe this as the private, home and recreational environments. Naturally, this will vary from student to student. Most students, however, spend a significant percentage of their waking hours involved in school-related activities. Classes start at 8:00 a.m. and do not break until 4:00 p.m. Most classes require out-of-class work and study, so evenings and weekends are often used, at least partially, for academic-related purposes. There is no question that at-home support, and individual tolerance for lost leisure time is a must for success in this environment.

(Implications: There is little we can do in planning and implementing the PLG to allow for specific immediate environment factors in the orienting context. However, it seems likely that, if the PLG were to save time, it would be welcome to the students.)

Organizational factors.

Perhaps the most salient orienting organizational factor of the VMRCVM is the curriculum employed, which would be considered a "traditional" curriculum of medical education. (The differences between traditional and problem-based medical curricula (the two predominant curricula used in medical education) have been discussed in the review of literature.) The Curriculum at the VMRCVM is a four-year, lock-step (meaning that all students take roughly the same core courses in the same order) curriculum. The first 3 years of the curriculum are made up largely of lecture-type classes with associated labs. Many courses teach and test memorization of verbal information and rules. Because the Clinical Pathology course occurs in the second semester of the second year, students prior to taking this course have had 3 semesters of a mostly lock-step, lecture-type curriculum. *(Implications: This contextual factor is one of the main impetuses for the creation of the PLG. The PLG is designed in part to make students deal with problems that cannot be solved through memorization of verbal information and simple rules.)*

Instructional Context

Learner factors.

Two major learner factors in the instructional context are how the learners perceive their role(s) and how they perceive their task(s) (Tessmer & Richey, 1997). The instructional context for the Clinical Pathology (VM8418) course already requires an alteration in how learners perceive both their roles and their tasks. This will be further changed after the implementation of the PLG and the related tutorials.

VM 8414 at Virginia Tech is taught in the context of cases. The bulk of each class period consists of presentations of case solutions by students, who have prepared previously with a professor. Each student is only required to present one case to the class, but all students prepare 2-3 cases for each class period, and there are 3 class periods per week. Pop quizzes are given throughout the semester -- averaging about one quiz every other class period. Students are required to work in groups of two. They prepare and present their cases in pairs and sometimes take their quizzes in pairs as well. While each student prepares all of the cases (with a partner), only one student presents any given case to the class at a time. The professor is able to give personal feedback to each student only once -- for the case they present in front of the class. While all students are still required to analyze each case, there is no system in place to give individual credit or feedback for their efforts.

This scenario requires several likely alterations in the way most learners perceive their roles and tasks. First, students must, while retaining the role of listener and note-taker, also assume, at times the role of teacher/lecturer. In terms of the learning task,

students must continue to learn and know verbal information and rules, but they must also engage in the higher-level task of synthesizing and explaining. While this change in role might be welcomed by some learners, it has been seen to be threatening/challenging to others.

(Implications: Some of the learner roles and tasks implicit in the Clinical Pathology course are new and different for many learners.)

Immediate factors.

VM 8414 is taught in a large lecture room capable of holding an entire class (90 students). Curved tables with attached chairs are situated in an ascending semi-circle facing the instructor area which is at the bottom, front of the room. At the instructor area are a large podium and projection screen. An audio system is used to make the teacher audible to the students. Students frequently complain that the chairs are uncomfortable. They are not padded, and they are anchored to the tables on hinges such that to sit down, one must swing the chair back away from the table. Gravity tends to carry the chairs towards the tables so that the students must keep at least one foot planted on the ground or a hand braced on the table to keep from slowly swinging stomach-first into the table. The lighting does not allow the projection screen to be visible at the same time that the room is adequately illuminated for reading and writing.

The instructor of VM 8414 acts not only as a lecturer, but also as a facilitator, requiring each student to meet individually with her at least once to discuss a case, and then to present that case in class. She also regularly assesses student learning through pop quizzes.

(Implications: Because the instructional setting is not a particularly comfortable one, if students can do what they would ordinarily do there in more comfortable settings (which the PLG and related tutorials make possible) the effect should be positive. Care must be taken not to endanger the positive aspects of the instructor's role as facilitator/mentor.)

Organizational factors.

For the purpose of this project, organizational factors are favorable. Grant money has been available to buy out professor time and to provide equipment, personnel, and software. A lab is available for students to use the PLG in. A department server is available, run by a competent and dedicated web master.

(Implications: We need to take advantage of these resources while they're available.)

Transfer Context

Learner factors.

Students are highly motivated to transfer the skills gained in VM 8414 to other settings. Most plan to become practitioners, and realize that lives will depend on their ability to accurately interpret laboratory data. Students will also be expected to transfer these skills to the clinical experiences if their fourth year.

(Implications: Students should be helped to understand how they will use the skills learned here in future environments.)

Immediate environment factors.

The context of the fourth year clinical experience as well as the context for the practitioner are quite unlike the learning setting. In both cases, the situations become more complex in many ways, involving real patients with real problems, real clients with their idiosyncrasies, and, in some cases other outside worries such as financial management.

(Implication: Efforts should be made to provide the students with as many similarities as possible in the learning environment to the transfer environment.

Organizational factors.

Because students will find themselves in a variety of specific transfer environments with varying value systems, organizational frameworks, etc., it is impossible here to plan for all possible future organizational factors.

Decision to Use a Web-Based Computer-Mediated Solution (Media Selection)

A web-based solution is required in this instance for several reasons. First, the funding for the project is contingent upon it. Second, a computerized solution is the only feasible way for the professor to be reasonably sure that all the students are completing the problem lists and receiving credit for doing so. She does not have time even to look at as many as eight hundred and ten cases per week (9 per student) for the purpose of assigning credit. Through a computer-based system integrated with a database, the cases will all be self-evaluated, and at a glance the professors will know how each student is doing, or will be able to look at and evaluate any given problem list.

Instructional Goal

Each learner will be able to 1) accurately and consistently identify and describe how pathophysiologic mechanisms cause data changes, and 2) group related data abnormalities under each mechanism. This process generally produces a diagnosis.

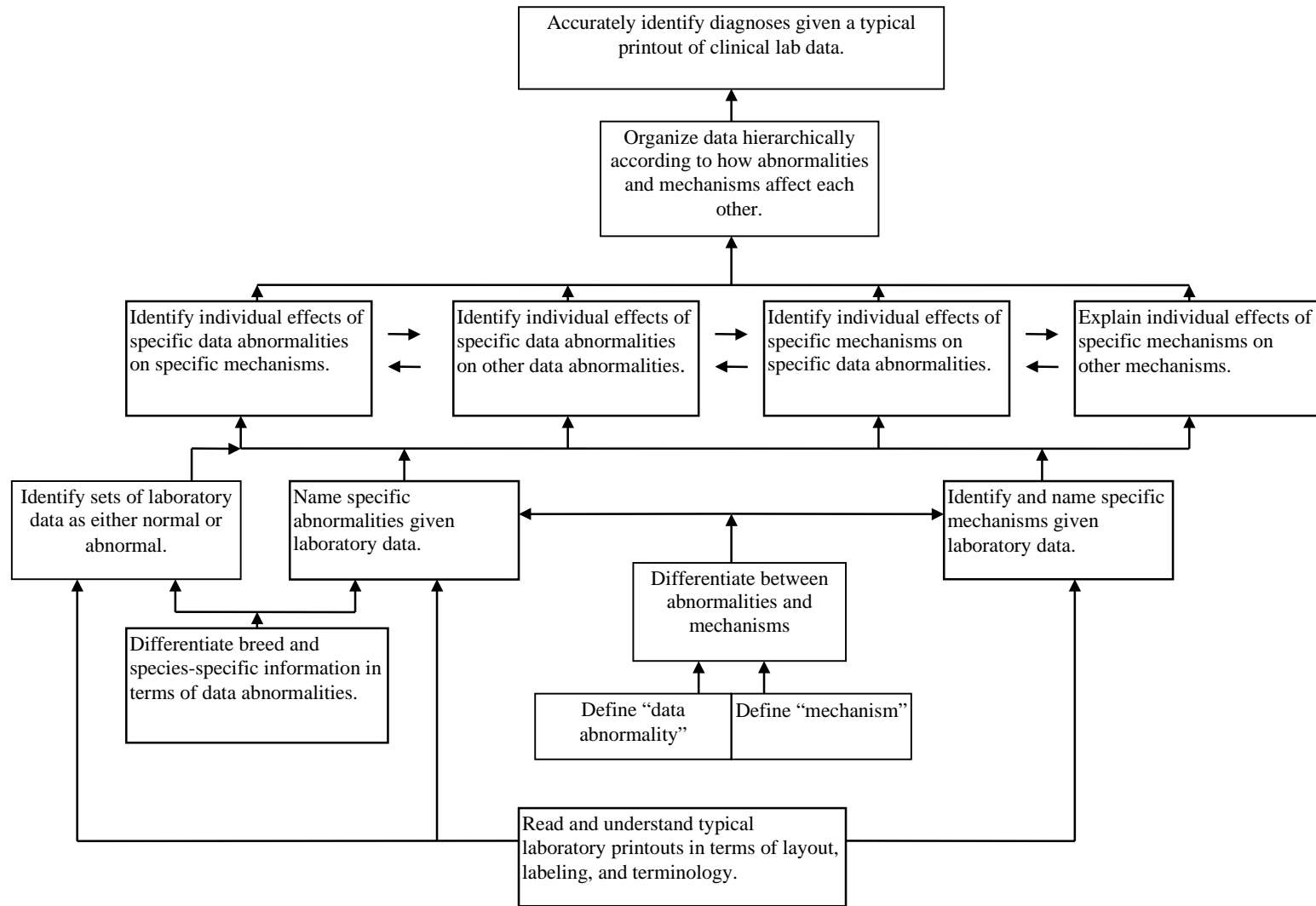
Instructional Analysis (See Figure B.1: Instructional Analysis)

The diagram in Figure B.1 is designed to illustrate, hierarchically, the skills required to arrive at the Instructional Goal. This diagram includes very few prerequisite skills for the course, however, some of the skills might turn out to be prerequisite to successfully using the PLG.

Performance Objectives

1. Accurately identify diagnoses given a typical printout of clinical lab data.
2. Organize data hierarchically according to how abnormalities and mechanisms affect each other.
3. Identify individual effects of specific data abnormalities on specific mechanisms.
4. Identify individual effects of specific data abnormalities on other data abnormalities.
5. Identify individual effects of specific mechanisms on specific data abnormalities.
6. Explain individual effects of specific mechanisms on other mechanisms.
7. Name specific abnormalities given laboratory data.
8. Identify sets of laboratory data as either normal or abnormal.
9. Differentiate between abnormalities and mechanisms.
10. Identify and name specific mechanisms given laboratory data.
11. Read and understand typical laboratory printouts in terms of layout, labeling, and terminology.
12. Differentiate breed and species-specific information in terms of data abnormalities.
13. Define “data abnormality”
14. Define “mechanism”

Figure B.1. Instructional Analysis



Instructional Strategy

It is customary, and generally logical to present the assessments and assessment instruments before presenting the strategy. However, in this case, since the assessments are built into the software and to gain an understanding of the software one must be familiar with the instructional strategy, I will present the strategy and then discuss assessment.

Literature Review

An in-depth review of literature (see the preliminary exam) led me to the following conclusions about how the PLG should be built, and therefore, inform the instructional strategy of the PLG. In essence, if the process being learned is that of dealing with clinical lab data, and all of the instructional objectives and assessments have to do with the learners' ability to do so, then it makes sense to have a strategy in which students are required to deal with clinical lab data. This is the essence of points 1, 2, and 5 of the following list which emerged from the review of literature:

1. Exposure to and practice dealing with medical problems should prove useful in the development of data classification skills, and, therefore the practice of such skills should be a part of the PLG.
2. Any teaching of diagnostic problem solving cannot rely on a forward-thinking, or "goodness of fit" model for dealing with problems, but should rather rely on a model which forces students to deal with the process of synthesizing content.
3. A tool such as the PLG should not be designed or implemented without allowing for context.
4. A tool such as the PLG should be built on a thorough front-end analysis.
5. A tool such as the PLG will allow time for learning and practice.

Step-by-step Explanation of the Instructional Strategy

The second major section of this document includes the full text of the Interface Design Document, which describes in detail how the learners will interact with the PLG. A full understanding of the interface is not necessary to understand the instructional strategy, however. I will present here those screens and interactions which embody the instructional strategy without the technical, precise, and laborious interface information available in the next section.

Each set of clinical laboratory data that the students will deal with will be embodied in a case -- a description of an animal with a certain set of symptoms and laboratory data. This use of cases is consistent with the course professor's current teaching method and is consistent with the idea from the review of literature that students should deal, as much as possible, with "real" problems.

Figure B.2, below is the Case Summary screen, which will allow students to learn initial information about the animal to which the case pertains. As can be seen, an image

of the patient is accompanied by a brief textual history. This screen is designed to provide the student with the essential information about the animal that would not be found in the laboratory data. Additional images and video can be added to accompany the case history if such images and video are considered instructionally useful for any given case.

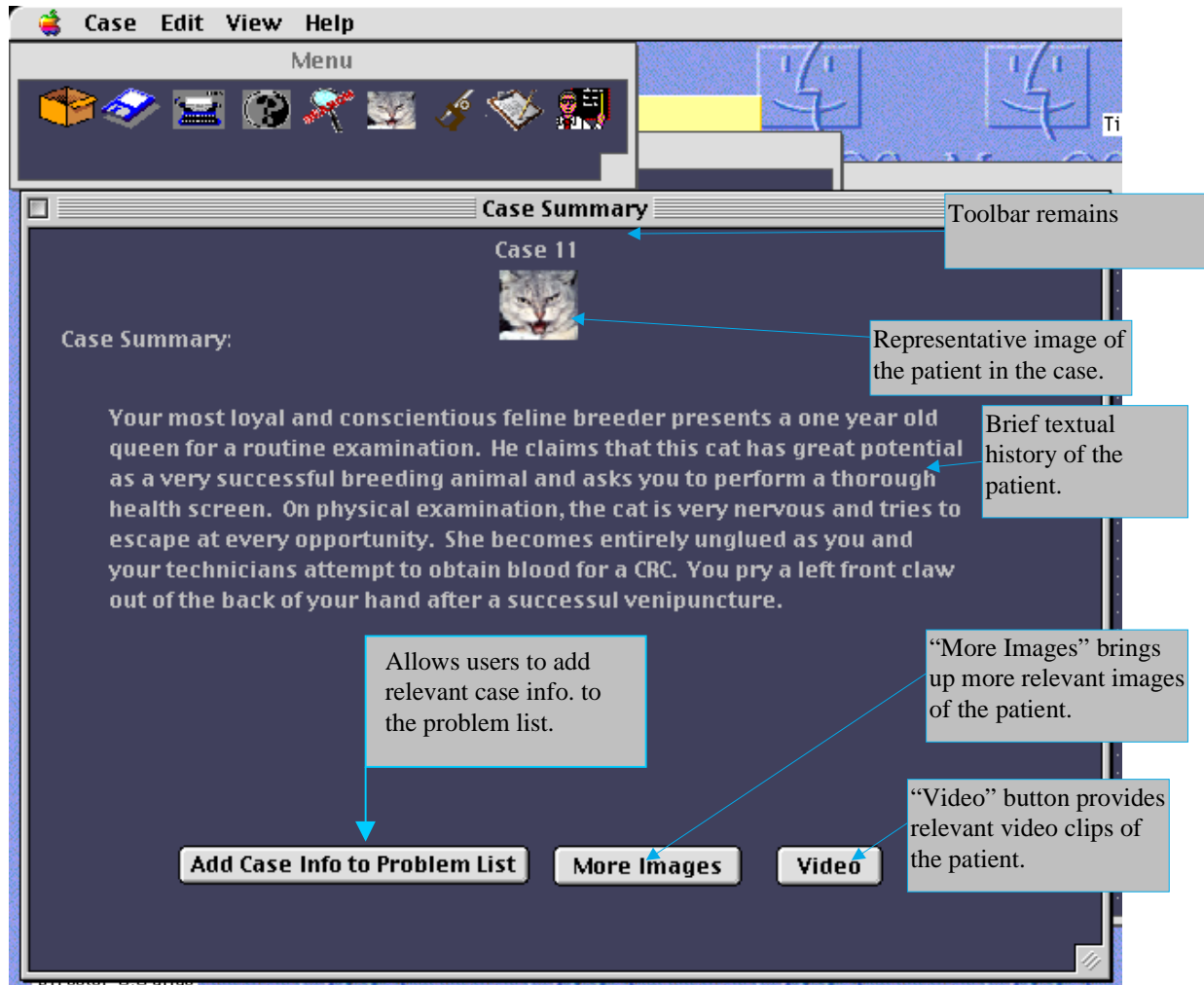
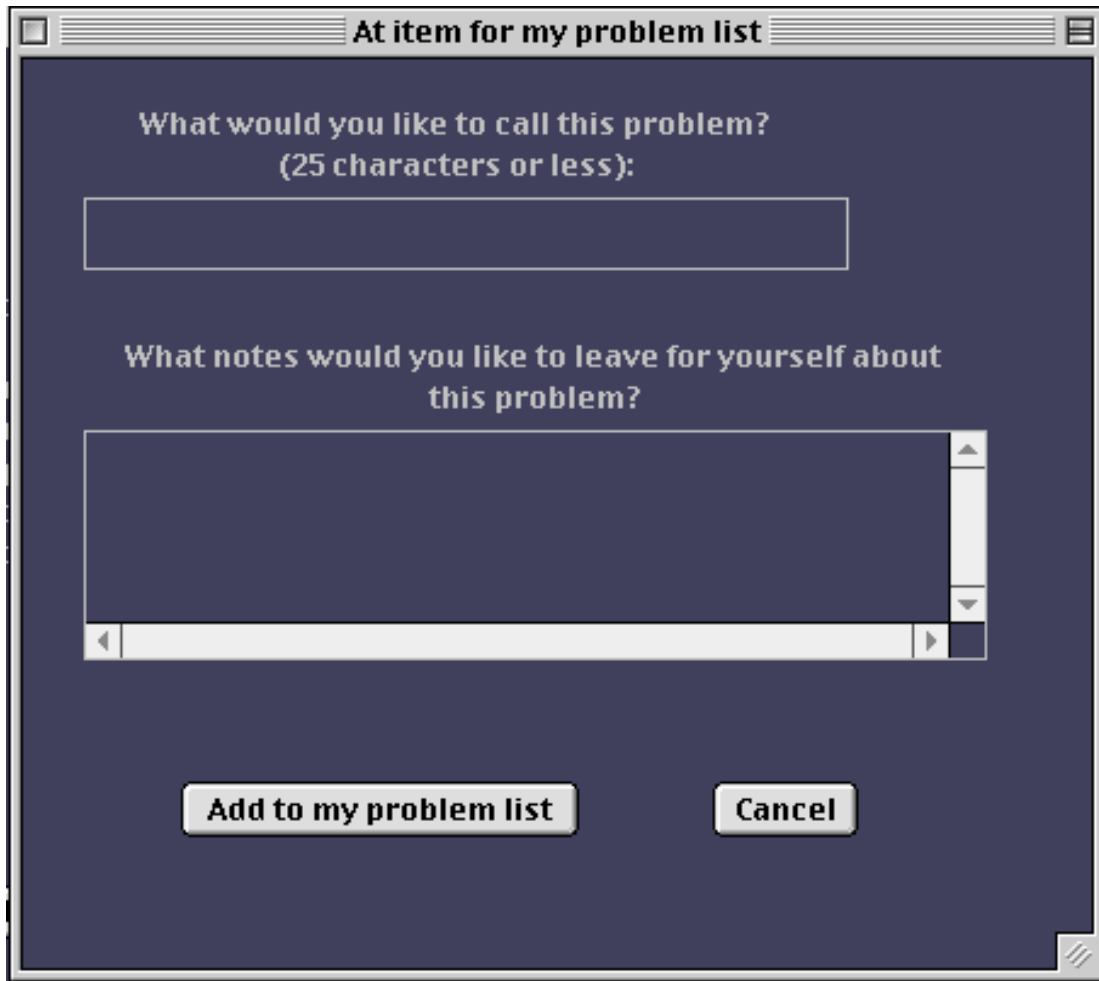


Figure B.2. Case Summary

When learners identify information in the Case Summary that they feel will prove useful to them in generating a problem list, but will not be likely to appear in the lab data, they will be able to enter that information by using the "Add Case Info to Problem List" button. Doing so will cause the screen shown in Figure B.3, below to appear. Once the user has entered a name for the problem or phenomenon identified, and has written a description of it (if desired), clicking the "Add to my problem List" button will result in the problem appearing in the "Data Abnormalities I've found" column of "My Problem List" (Figure B.5).



The image shows a dialog box window with a title bar that reads "At item for my problem list". The window has a dark blue background. The first question is "What would you like to call this problem?" followed by "(25 characters or less):". Below this is a single-line text input field. The second question is "What notes would you like to leave for yourself about this problem?". Below this is a multi-line text area with a vertical scrollbar on the right and a horizontal scrollbar at the bottom. At the bottom of the dialog box are two buttons: "Add to my problem list" and "Cancel".

Figure B.3. Adding an item to the problem list from the lab data screen

The next interaction of instructional consequence that will be available to the learners is illustrated in the data classification screen. Here the learners initially encounter the laboratory data. This screen and can be seen in Figure B.4, below.

The data classification screen is designed to allow students to classify each item of laboratory data as normal, false normal, high, false high, low, false low, or abnormal. This is accomplished by manipulating the small pull-down menus appearing to the right of each data point. If data points are changed to anything other than normal, a text box

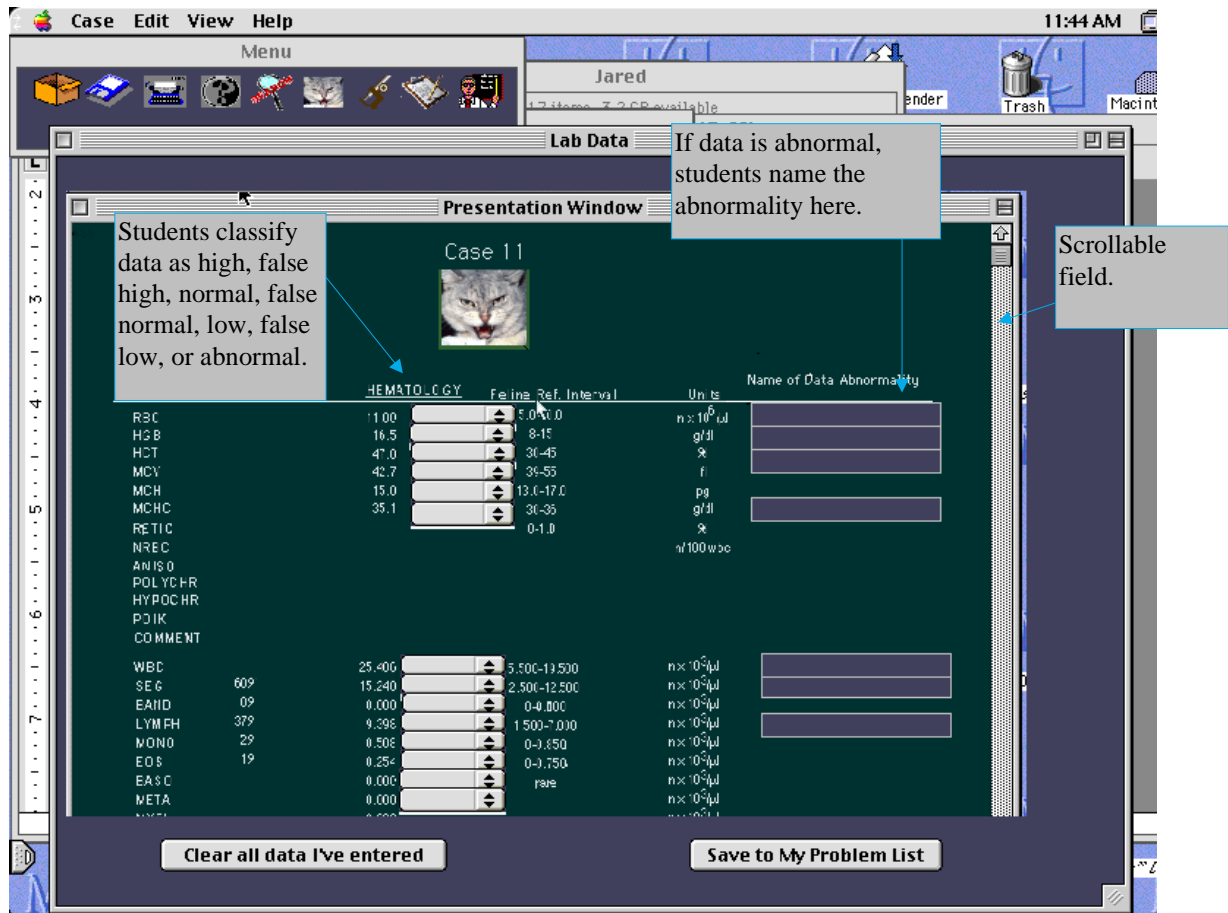


Figure B.4. Lab Data

such as the ones shown in the right column of Figure B.4 will appear. In each text box the learner will enter the name of the corresponding data abnormality.

If a data abnormality name is entered which does not match the appropriate name in the database, a dialogue box will appear indicating that the name entered is not correct. At the third unsuccessful attempt to name the data abnormality, the same dialogue box will appear, only this time, it will also include a button labeled “Tell me.” If the “Tell me” button is clicked, the dialogue box will disappear, and another will appear stating “When [name of data] is [high or low] it is known as [name of data abnormality].” Checking the students at this point is important for several reasons. First, if the students proceed to the problem list with incorrectly identified data abnormalities, such a mistake would probably lead to problems in generating an accurate problem list. Also, since knowing the names of data abnormalities is an instructional objective, and since it is feasible for the system to check this knowledge, it seems logical to do so at this point. Students are given the correct name of the data abnormality after three unsuccessful attempts, on the one hand to reduce the likelihood that they will simply allow the system to do the work of naming the data abnormalities for them, and on the other hand to make sure that they are not prevented from completing the exercise for lack of remembering how a name is spelled.

The lab data screen is purposefully laid out to reflect the printouts that veterinary practitioners are forced to work from in the field. Learning theory suggests that the more a practice environment approximates the environment in which the task will have to be performed, the more likely the knowledge and skills gained are to transfer to the work environment.

The system will not allow learners to begin a problem list until all identified data abnormalities are named correctly. This will ensure that the learner does not begin a problem list without all the necessary pieces for completing it (as identified by that learner).

Once the learner has identified and named all the relevant data abnormalities, the problem list screen, as shown in Figure B.5 will become available. Any time a data abnormality is named (Figure B.4), or a problem is added to the problem list from the screen shown in Figure B.3, the data abnormality or problem identified appears in the column on the right hand side the "my problem list" screen (Figure B.5). Students create a problem list by dragging data abnormalities from the column on the right (Figure B.5) to the box on the left and manipulating them there. In the box to the left (Figure B.5) students can enter mechanisms, re-label abnormalities as mechanisms (if they function as such), add notes in text boxes attached to each mechanism or abnormality, and arrange mechanisms and abnormalities in relation to each other. The problem list communicates causal relationships between items of data which are arranged in an outline format. That which is above and to the left causes that which is below and to the right. Abnormalities on the same level are both caused by the same thing, but do not affect each other.

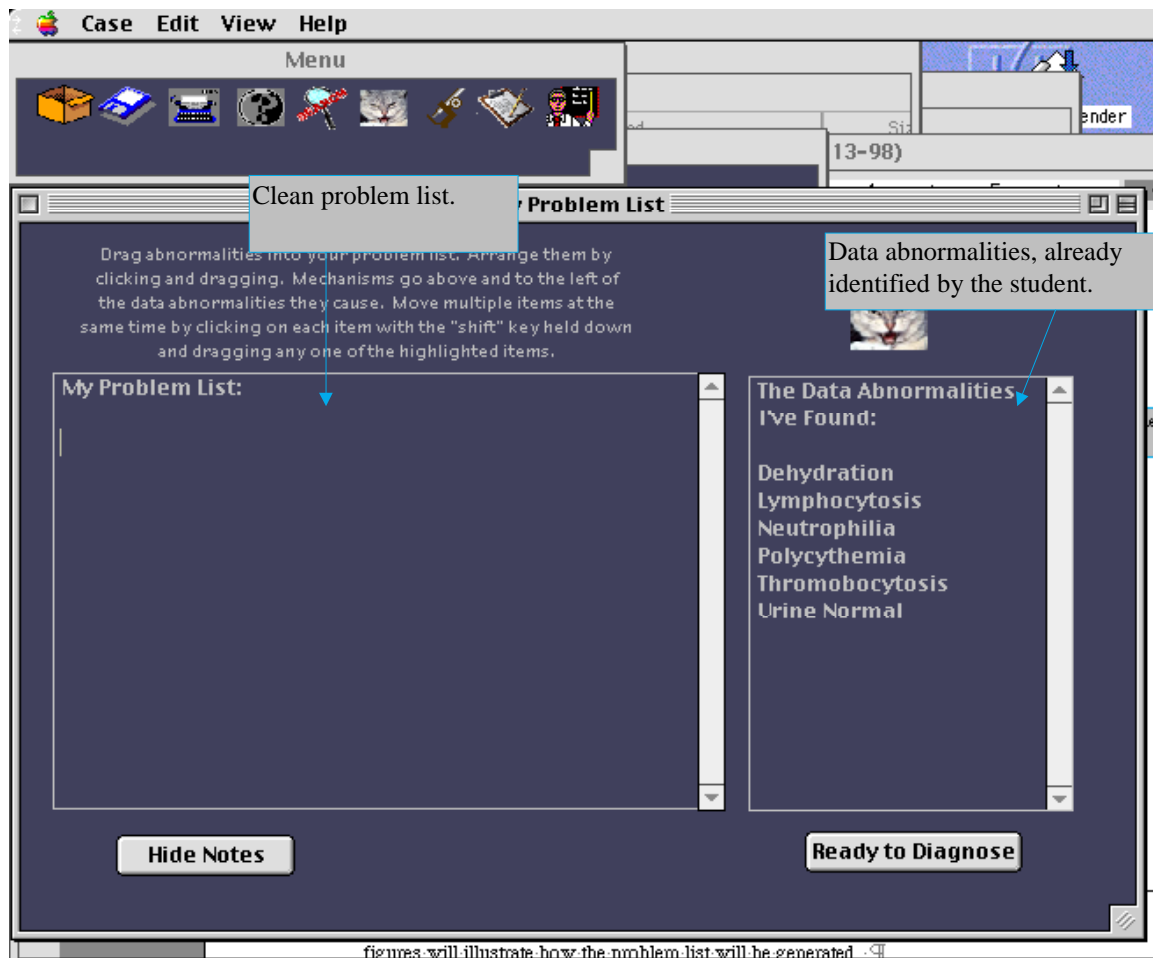


Figure B.5. The User's Problem List

Figure B.6 (below) illustrates a completed problem list. This problem list communicates (among other things) that the mechanism "Splenic Contraction" causes polycythemia, and thrombocytosis. The notes fields for "difficult phlebotomy" and "cat is struggling" are visible while the other notes fields are not. This screen is designed as a "scaffold" of sorts for the learners. It allows them to see at once all the data abnormalities they have identified and requires them to organize them all in a meaningful way (one which identifies causal relationships.) Thus, this screen, while not providing the answers, is designed to force the learners, in a way, to ask all the right questions.

The three screens explored up to this point are all available to the learner throughout the process of generating the problem list. This is an iterative process. Students might decide that what had appeared to be "high" data, is actually "false high", etc., and will be able to go back and change it in all relevant screens accordingly.

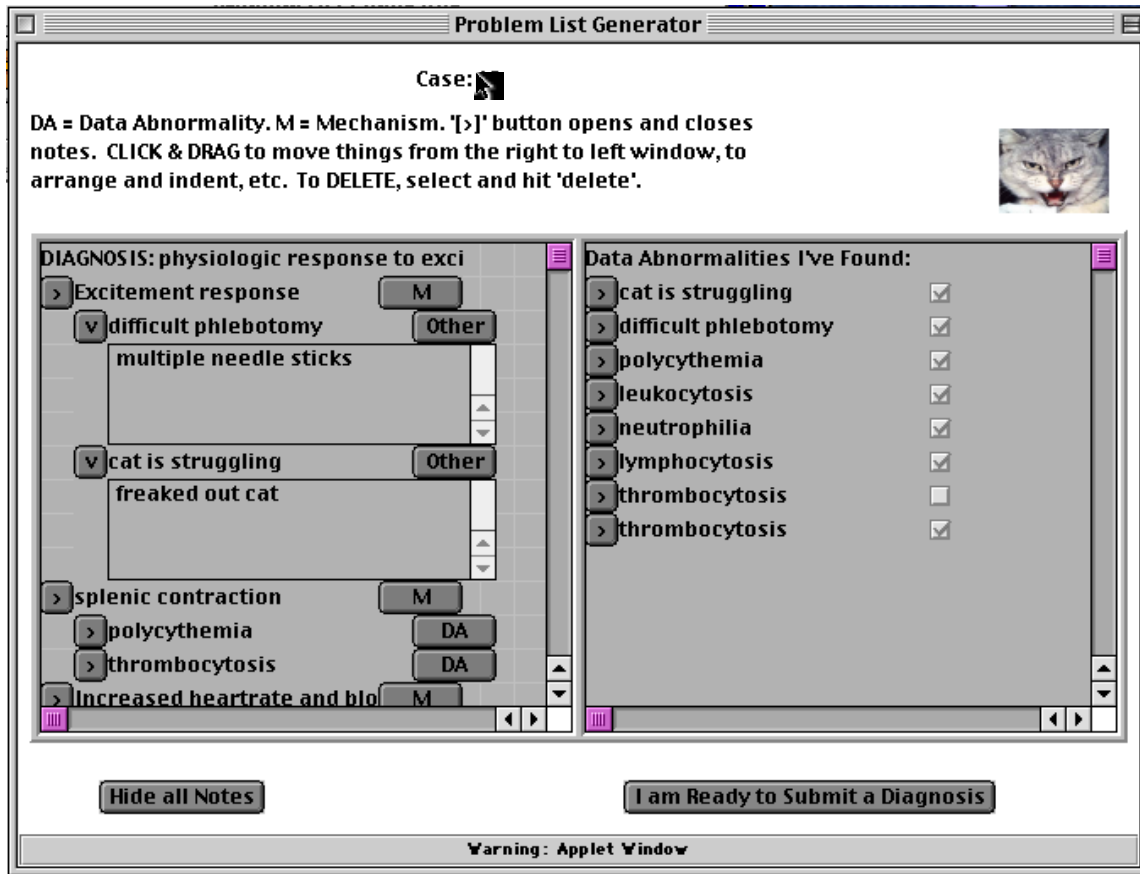


Figure B.6. Completed Problem List

When students have created a problem list to their satisfaction, they can submit a diagnosis and compare their problem list to the expert problem list. To prevent the students from simply copying the expert problem list instead of going through the process of generating their own, their problem list will no longer be able to be modifiable once they have submitted their diagnosis (which makes the expert problem list available).

The expert problem list can be seen in Figure B.7, below. As shown, this screen also displays the learner's problem list so that the two may be compared. This screen is designed to answer the questions that learners might not have been able to answer while generating their own problem lists, and to reinforce for learners that which they had identified correctly. Here the learners can read the expert solution, complete with notes, and compare their solution with the expert solution. This screen is designed to provide the learners with meaningful feedback on their efforts, fill in knowledge gaps that may have been missing, and prepare the learner for the lecture in which the professor will answer questions on this problem. (Learners will be able to print their problem lists or the expert problem lists for study or review purposes.)

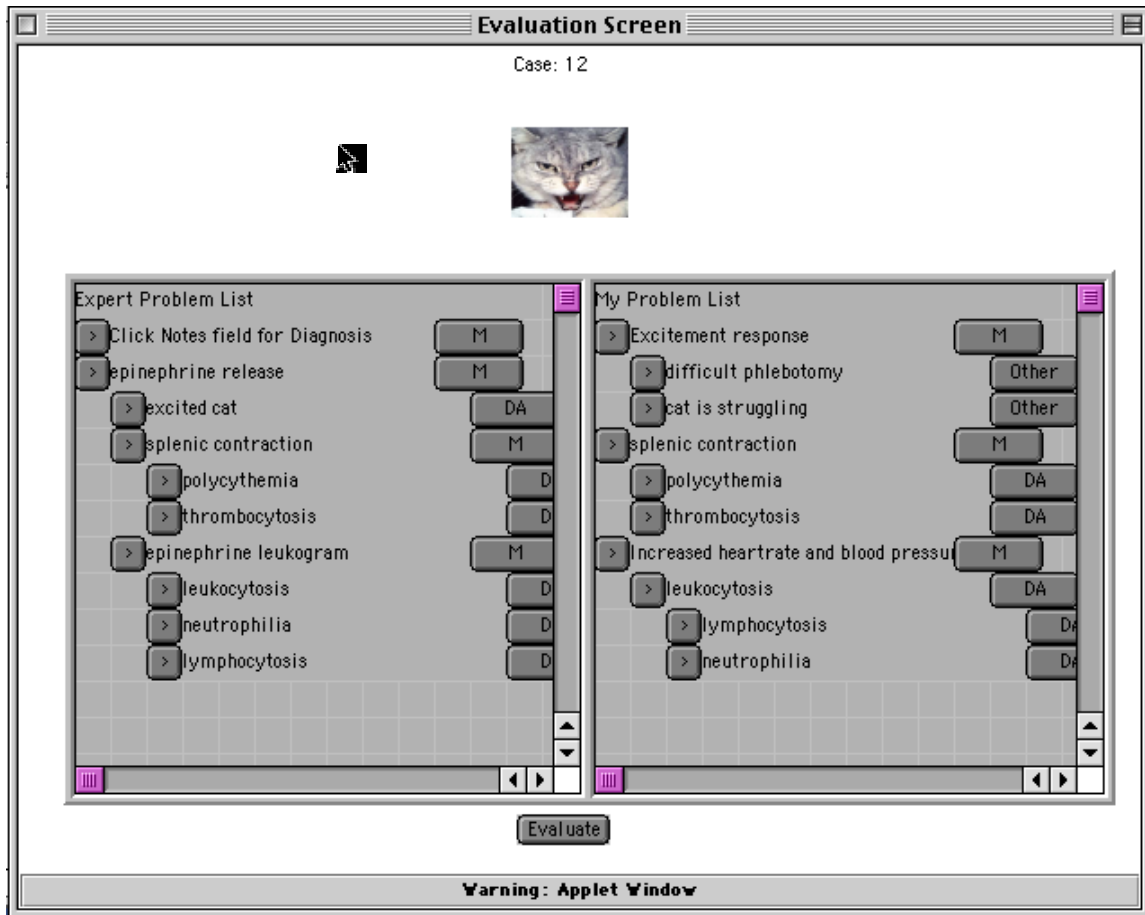


Figure B.7. "My Problem List" and "Expert Problem List" compared

As seen in Figure B.8, once they have seen the expert problem list, learners will evaluate their own problem list on a scale of 1 to 5 for the purpose of reporting how closely it matches the expert problem list. This self-evaluation will not be used for grading purposes, but will be available to the professor for an indication of how each student feels he/she is doing.

Summary of the Instructional Strategy

The strategy described here might be seen as what Mandin , Jones, Woloschuk, and Harasym (Mandin, Jones, Woloschuk, & Harasym, 1997) call providing students with an "expert scheme" and problems to apply the scheme to. In this case, the problems are cases which increase in length and complexity as the semester progresses. The expert scheme requires students to deal with the data in sizeable chunks, first by viewing a case history, then by classifying data and identifying and naming data abnormalities, and finally by arranging all the aberrant data in a causal structure. The strategy also allows students to compare their solutions with an expert solution.

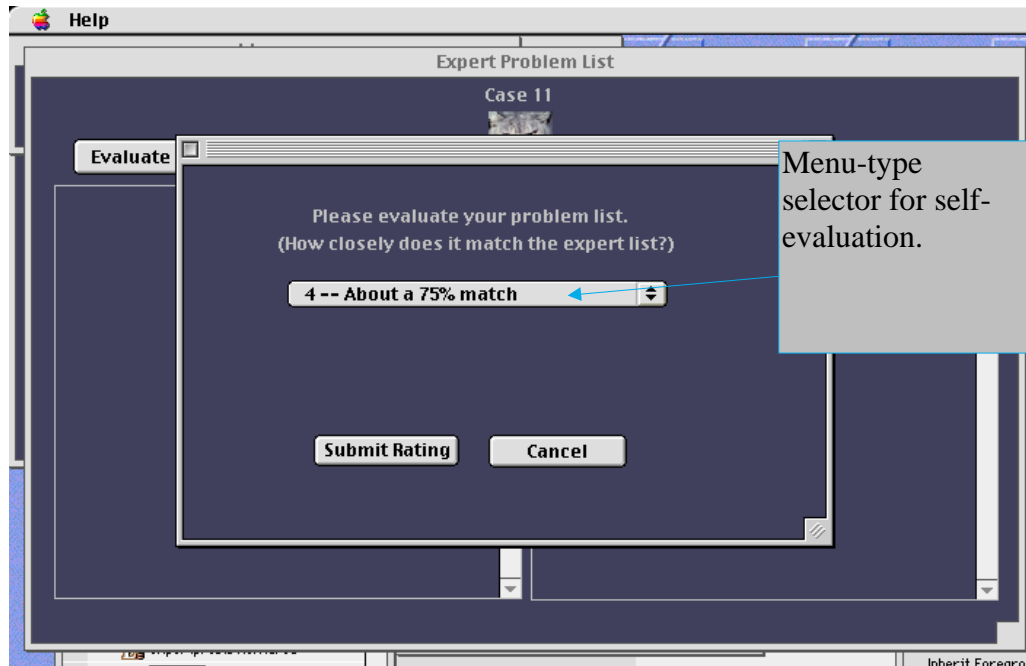


Figure B.8. Evaluation Dialogue

Assessment Plan

The process of generating a problem list is a complicated one. In many cases, more than one diagnosis might adequately explain a given set of laboratory data. Furthermore, it is felt that the process of arriving at a reasonable diagnosis is, in this case, more important than arriving at an exact diagnosis. This process can be evaluated at two points. First, it can be determined whether or not students have correctly identified and named any given data abnormality. Second, the finished problem list itself is an artifact which reflects the students' understanding of the data that led to it. Therefore, students will be assessed at those two points. The following will describe how the PLG will deal with student assessment for each performance objective. The performance objectives can be referenced by number to the performance objectives section appearing previously.

1. Objectives 8, 11, and 12 have to do with recognizing and correctly classifying data abnormalities. Because of the inter-reactions of abnormalities and mechanisms, in some cases data might appear abnormal but actually be normal, or visa versa. Therefore, automated feedback will not be given as students classify the data, and students will be allowed to change their data classification at any time. Also, students will be allowed to note why they classify data as they do.
2. Objectives 9, 13, and 14 have to do with the learners' understanding of what mechanisms and abnormalities are, and is prerequisite to correctly organizing mechanisms and abnormalities. Students will not be directly evaluated on these objectives, but their understanding of these principles will be demonstrated in the way they organize abnormalities and mechanisms in the PLG.

3. Objective 10 has to do with learners' ability to name specific mechanisms. Once students have identified a mechanism, they will be required to enter its name. Because of the wide variety of possible names for mechanisms, the system will not check the name, but students will auto-evaluate when presented with the expert problem list.
4. Objective 7 has to do with learners' ability to name specific data abnormalities. Once students have identified an abnormality, they will be required to enter its name. The program will do a check to verify whether or not the data abnormality has been identified accurately, and will provide immediate feedback.
5. Objectives 3, 4, 5, and 6 are all interrelated. Meeting these objectives leads to objective 2 being met, which in turn leads to objective 1 being met. Because of the complexity of organizing the data accurately, and because the data can be organized correctly in more than one way, students will be required to organize all the data and arrive at a diagnosis without receiving feedback on their overall solution. Once students feel that they have accurately diagnosed the problem, they will be able to receive feedback in the form of an expert solution to the problem. The expert solution will consist of the "problem list" (basically, the end result of performing this entire process) as performed by an expert. Students will then compare their problem list with the expert problem list, and submit an auto-evaluation (1-5) of their work as it compares to the expert solution (5 being a perfect comparison and 1 being a solution which has nothing in common with the expert solution). This auto-evaluation will provide the professor with a quick way to gauge each student's progress. The act of comparing their solutions with the expert's will allow students to correct misconceptions or reinforce knowledge they understand. The auto-evaluation will not be used for grading (other than to give students credit for having completed the assignment.) The students' solutions will be preserved so that students can review their work, or professors can peruse particular students' solutions to determine, more fully, how they are doing in class.

Evaluation Plan

The principle purpose of the evaluation is to improve the PLG, and to verify its overall clarity, impact and feasibility. The first three steps described below are designed to serve both purposes, while the latter is more of a summative tool which will aid in making general comparative claims about the overall usefulness of the PLG. The first three elements of this evaluation plan are based on Dick and Carey's (Dick & Carey, 1996) plan for conducting a formative evaluation. The fourth element employs a time-series evaluation design for making some quantitative claims for the overall effectiveness of the PLG.

1. Expert Review

In this process the functional prototype and instructional and interface designs will be given to at least one instructional expert, one subject matter expert, and one interface design expert for review and suggestions.

Instructional Expert: -- Dr. Barbara Lockee and Dr. John Burton will serve as instructional experts for the project.

Subject Matter Expert: -- Dr. Holly Bender and will serve as the Subject Matter Expert. (Several other experts in the field of clinical pathology may also review the PLG.)

HCI Expert: -- Dr. Deborah Hix, professor of computer science and Jay Swami, practicing interface designer will provide expert feedback on the interface.

2.One-to-one Evaluations

Participants

The one-to-one evaluation will be carried out with at least three learners, one considered advanced, one considered average, and one considered below average. In this context, advanced, average and below average refer to the student's ability to classify and interpret lab data, and not to the student's class ranking or performance in previous classes. Students will be evaluated and categorized by Dr. Holly Bender in an interview setting where she and each student discuss a specific case. These interviews are a regular part of the course. Participants in the one-to-one sessions will be selected from a pool of volunteers who have been evaluated by Dr. Bender.

Observation

Students participating in the one-to-ones will be observed and videotaped. Dick and Carey's three categories, "clarity", "impact" and "feasibility" will be used as general categories for observation during these one-to-one sessions.

Clarity.

Because the PLG is to be computer-mediated, the interface is certainly a major factor in the "clarity" equation. (The clarity of new information presented will not need to be evaluated since the PLG does not present new information). It seems clear that if the interface is not usable, the PLG will not be successful. Therefore, for evaluating clarity, I will adopt and borrow a procedure designed by Hix and Hartson (Hix & Hartson, 1993; Hix & Hartson, 1997) for testing interfaces. This procedure involves identifying and testing "usability specifications." In this process, the evaluator identifies key "benchmark" tasks that are designed to test every function of the system. Each participant is observed performing each benchmark task and one or more attributes such as time taken to complete a certain task and number of mistakes for first time use are recorded. Descriptive data is also recorded, particularly when learners encounter "critical incidents" -- moments at which it is difficult or impossible for the learners to know what to do next. Quantitative results are recorded on a "Usability Specifications" table (see Figure B.9 for an example of a usability specifications table with some usability specifications I intend to measure for the PLG), and other observations are recorded on a separate data collection sheet (see Figure B.10) and mapped to the appropriate tasks. In this particular context, Hix and Hartson's method will be altered slightly. Time on-task will be recorded,

but time "manipulating" the interface will be separated from time spent dealing with the instructional tasks of classifying lab data or generating a problem list. Also critical "learning" incidents (moments in which the learners are stumped by the learning problem) will be separated from interface-related critical incidents.

Figure B.9. Usability Specifications * (Adapted from Hix and Hartson (Hix & Hartson, 1997))

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	Observed Results -- interface	Observed Results -- instruction
#1 Initial Performance	“Get into the system” per benchmark 1	Length of time taken to successfully complete task				
#2 Initial Performance	“Open Case 11” per benchmark 2	Length of time taken to successfully open Case 11				
#3 Initial Performance	“Print My problem List” per benchmark 3	Length of time taken to successfully complete the task				
#4 Initial Performance	“Find out” per benchmark 4	Length of time taken to successfully complete the task				
#5 Initial Performance	“Find, mark and label data” per benchmark 5	Length of Time taken to successfully complete the task				
#6 Initial Performance	“Create a problem list” per benchmark 6	Length of Time taken to successfully complete the task				

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	Observed Results -- interface	Observed Results -- instruction
#7 Initial Performance	“Create a problem list” per benchmark 6	Number of errors observed while completing the task				
#8 Learnability (after 10 minutes of practice)	“Create a problem list” per benchmark 6	Length of time taken to successfully complete the task				
#9 Learnability (after 10 minutes of practice)	“Create a problem list” per benchmark 6	Number of errors observed while completing the task				
#10 Initial Performance	“Create a note” per benchmark 7	Number of errors observed while completing the task				
#11 Initial Performance	“Create a note” per benchmark 7	Time taken to complete the task				
#12 Initial Performance	“Submit diagnosis” per benchmark 8	Time taken to complete the task				
#13 Initial Performance	“Report” per benchmark 9	Number of errors observed while completing the task				
#14 Initial Performance	“Report” per benchmark 9	Time taken to complete the task				
#15 Initial Performance	“See if word appears” per benchmark task 10	Number of Errors observed while completing the task				

USABILITY ATTRIBUTE	MEASURING INSTRUMENT	VALUE TO BE MEASURED	CURRENT LEVEL	TARGET LEVEL	Observed Results -- interface	Observed Results -- instruction
#16 Initial Performance	“See if word appears” per benchmark task 10	Time taken to complete the task				
#17 Initial Performance	“Close Case 11” per benchmark task 11	Time taken to complete the task.				
#18 Initial Satisfaction	Questionnaire	Average scores from all items on the Questionnaire. (Includes items #4, 6, 7 10, 11 and 13 of Quis Questionnaire)				
#20 First impression	User reaction	Number of negative/positive remarks during session				

*A description of each benchmark task may be found in Appendix B.1.

TASK NAME:	Participant ID: Task Start Time: Time to Perform Task:		Date: Task End Time:
No. of errors:			
Critical Incident Description	Tape Counter	Evaluators Comments	
1.			
2.			

Figure B.10. Critical Incident Evaluation Sheet

Learners will be encouraged to talk as they work with the PLG, and the observer will ask questions when it seems necessary, to understand what the participants are thinking or feeling.

Feasibility.

Results of the procedures for testing clarity will also prove useful for evaluating feasibility. Time taken to perform given tasks will probably be the best measure of feasibility. However, other factors will be noted such as how comfortable the learners are dealing with computers in general, how conducive the test environment (potentially a busy lab) is to completing the tasks, and how quickly the system responds. The observer will take care to notice any other feasibility problems/issues.

Impact.

Two immediate indicators of impact will be available. First, the problem list generated by the learners will be available and the accuracy/completeness of the problem list will be evaluated by a subject matter expert. Second, at the conclusion of the one-to-one session, the observer will ask each participant for general impressions of their experience using the PLG. Third, each participant will be asked to complete a brief questionnaire at the conclusion of the one-to-one sessions, including lichert-type indicators of their experience using the system and their confidence-level after using it. This questionnaire will be constructed following the expert review process, and will reflect potential problem spots identified by the experts.

3. Small Group Evaluation

Between 8 and 20 learners (based on recommendations by Dick and Carey (Dick & Carey, 1996) will participate in the Small Group Evaluation. Members of the small group will be given a case to complete, instead of benchmark tasks. They will not be timed or video taped as were the one-to-one participants. The observer will watch and note to what extent changes that emerged as a result of the one-to-one testing have been effective, and to identify any previously missed problems. Observations will be descriptive in nature and will seek to discover any major problems that may have been missed. The small group sessions will be followed by a small-group focus-group session in which students will be asked for their candid reactions to the PLG. This session will be video or audio taped. Students will also be given the same questionnaire that was administered to one-to-one participants, and their generated problem lists will also be evaluated by a subject matter expert.

4. Field Trial

A field trial will be conducted. It will be carried out in a real setting by roughly 30 students. In this case, a real setting is anywhere the students happen to find a computer with a java-enabled browser that they care to use. Students will receive a questionnaire to determine their individual experiences using the PLG. This questionnaire will not be

designed until after the one-to-one evaluations and the small group evaluations. Experience doing those evaluations will prove useful in designing an appropriate questionnaire. In general the questionnaire will contain items evaluating clarity, impact and feasibility, as was the case with the other two evaluations. Student's problem lists will also be analyzed by a subject matter expert.

5. General Data Collection and Analysis for Estimating Impact of the PLG

As was discussed in our pre-prospectus meeting, it is hoped that by gathering certain types of data from the entire class, it might be determined to what extent the PLG is helping/hurting certain groups of learners based on their ability levels. We also hope to acquire some data that will give us the ability to make some quantitative claims about the effects of the PLG. A longitudinal, single-group time-series design as described by Fitz-Gibbon and Morris (Fitz-Gibbon & Morris, 1978) will be used for this part of the evaluation, as described generally below.

During the first part of the semester, the PLG will not be used by the students except for in small groups as needed for evaluation purposes for the first three evaluation stages. Near the midpoint of the semester, the entire class will begin to use the PLG. Throughout the entire period of the course the students will receive several quizzes a week which generally test their ability to work through a case that has already been worked through in class and indicate either how well students have paid attention in class as the solutions to the cases have been presented, or how well they have understood the concepts, or both. Several Lichert-type attitudinal measures will be included with those quizzes. (See Figure B.12 below for an example quiz).

The quiz data will be watched, and any significant trend or level changes that coincide with the introduction of the PLG will be noted. The outcomes will not be analyzed statistically *, but will be reported as graphs, as described by Fitz-Gibbon and Morris (Fitz-Gibbon & Morris, 1978). Overall scores on these tests will also be analyzed by "ability group" using the categories of advanced, average, and below average decided upon by Dr. Bender. These group scores will be graphed to determine possible effects of the PLG on different groups of learners. Scores will also be grouped and analyzed according to those who participated in earlier stages of the evaluation and those who did not, to account for any differences in performance due to having used the PLG in some form or other before this point. (*Note, Fitz-Gibbon and Morris encourage against the researcher applying statistical tests to a single-group time series design because of the potential complexities involved and the questionable returns. They argue that graphing such data provides adequate information for conjecture about whether or not a certain treatment has been effective. However, there are ways of analyzing this data statistically, and if the committee feels that to do so would be beneficial, I would be more than happy to do it.)

Reporting

The methods and results of the evaluation will be written up and reported using standard APA format where applicable in the methods and results sections of the final dissertation document.

1. Characterize the neutrophilia and prognosis: (choose the **correct** answer for 1 pt.)
 - a. Mature neutrophilia, without a left shift, favorable prognosis
 - b. Neutrophilia, with clinically significant and degenerative left shift, unfavorable prognosis
 - c. Neutrophilia, with clinically insignificant and regenerative left shift, favorable prognosis
 - d. Neutrophilia, with clinically insignificant and degenerative left shift, favorable prognosis
 - e. a neutrophilia with a leukemoid left shift, unfavorable prognosis

2. If you were to perform a marrow aspirate, what would you expect to find in the neutrophil precursor pools? (choose the **2 correct** answers for 2 pts.)
 - a. a proliferation pool with normal cellularity
 - b. a depleted proliferation pool
 - c. a hyperplastic proliferation pool
 - d. the data does not imply anything about the proliferation pool
 - e. storage pool with normal cellularity
 - f. a depleted storage pool
 - g. a hyperplastic storage pool
 - h. the data does not imply anything about the storage pool

3. Choose the cause of the anemia: (choose the **correct** answer for 1 pt.)
 - a. acute blood loss, with no time for a regenerative response
 - b. extravascular hemolytic crisis
 - c. iron deficiency due to chronic blood loss
 - d. iron sequestration by macrophages
 - e. deficiency of erythropoietin

4. Which statement best describes the movement of neutrophils in this patient, compared to a healthy animal? (choose the **correct** answer for 1 pt.)
 - a. neutrophils are entering the blood from the marrow at a higher rate than normal, and entering the tissues at a higher rate than normal
 - b. neutrophils are entering the blood from the marrow at a lower rate than normal, but are entering the tissues at a higher rate than normal
 - c. neutrophils are entering the blood from the marrow at a lower rate than normal, and entering the tissues at a lower rate than normal
 - d. neutrophils are entering the blood from the marrow at a higher rate than normal, but are entering the tissues at a lower rate than normal
 - e. neutrophils are entering the blood from the marrow, and entering the tissues at a normal rate.

Please rate the following items using the following scale:

1	2	3	4	5	6
Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree

- | | |
|---|-------------|
| 5. I am confident that I will do well on the next set of problem lists. | 1 2 3 4 5 6 |
| 6. Dealing with clinical problems is a lot of work. | 1 2 3 4 5 6 |
| 7. I just do not understand Clinical Pathology. | 1 2 3 4 5 6 |
| 8. Doing the problem lists helps me understand Clinical Pathology. | 1 2 3 4 5 6 |

Figure B.12. Sample Quiz

Appendix B.1: Specific Benchmark Task Descriptions

Benchmark Task #1

Task Statement: Get into the system. Use the name “George” and the password “jungle”.

Objective: To verify that the way to logon to the system is clear.

Steps involved:

At the Logon (introductory) screen:

- Type “George” into the name field
- Return, tab, or mouseclick to get to the password field.
- Type “jungle” into the password field.
- Click the “Logon” button.

Starting point: Introductory screen.

Ending point: Blank screen with menu and toolbar.

Possible errors:

- User fails to enter the name or password correctly.
- User fails to click the “logon” button.

Benchmark Task #2

Task Statement: Go into Case 11. You will know you are there when you see a screen titled “Case 11” and subtitled “Case Summary

Objective: To verify that the way to open a case is clear.

Steps involved:

- Open the Case Menu
- Select the open option
- Select Case 11 in the pull down menu
- Click the Open button

or

- Select the Open button from the menu bar
- Select Case 11 in the pull-down menu
- Click the Open button

Starting point: Introductory screen after log-in. Time measurement starts from the moment instructions provided are read by the user.

Ending point: Case 11 Case Summary Screen. Time measurement ends when user declares he/she has opened case 11.

Possible errors:

- User fails to use the pull-down menu
- User selects the wrong option in the case menu
- User fails to select the correct case in the pull-down menu
- User fails to click the “Open” button

Benchmark Task #3

Task Statement: Assuming that it has already been created, Print “my problem list” for Case 11 with all notes.

Objective: To test the usability of the printer interface.

Steps involved:

- Open the case menu (unless using the toolbar)
- Select the print option from the case menu or the printer button on the menu bar
- Select the printer (if necessary)
- Check the “My Problem List” box
- Check the “Print all notes” box
- Click the print button

Starting point: My Problem List screen. Time measurement starts from the moment instructions provided are read by the user.

Ending point: Print dialogue box. Time measurement ends when the user clicks the “print” button.

Possible errors:

- User fails to find the print item in the Case menu, or on the toolbar.
- User fails to specify the correct printer.
- User fails to check appropriate box and radio button in the print dialogue box.
- User fails to press the print button.

Benchmark Task #4

Task Statement: Find out whether or not the cat in Case 11 clawed the examining veterinarian during the routine physical exam. Once you know, tell the observer.

Objective: This benchmark task is to verify whether or not it is clear that general case information would be available in the “Case Summary”, and that the “case summary” is easy to access and read.

Steps involved:

- Open the “View” menu (unless using the toolbar).
- Select the “Case Summary” menu item from the menu, or the “Case Summary” icon on the toolbar.
- Read the brief textual description provided under the picture.

Starting point: Lab data screen (or any other screen, once case 11 has been opened, other than the case summary screen). Time measurement starts as soon as instructions are given and user has verified that he/she understands the task.

Ending point: User tells the observer that the cat did indeed claw the veterinarian.

Possible errors:

- User fails to recognize that the required data is found on the “Case Summary” screen.
- User fails to find the “Case Summary” screen.
- User fails to read the info. in the “Case Summary” screen adequately to answer the question.

Benchmark Task #5

Task Statement: Find all data abnormalities in Case 11. Mark these abnormalities appropriately, and label them by name.

Objective: This benchmark task is to verify that the “lab data” form can be found and used by the users.

Steps involved:

- Open the “View” menu (unless using the toolbar).
- Select the “lab data” menu item from the menu, or the “lab data” icon on the toolbar.
- Identify 3 pieces of data that are abnormal (including polycythemia).
- Using the small pull-down menu with “H”, “L”, and “N” items to the right of each piece of information, label each as high (H), or low (L). (There are probably no “N”s, as the task is to find abnormal data.)
- Enter the name of each data abnormality in the text box to the right of each data abnormality.

Starting point: Case 11 “Case Summary” screen. Time measurement starts as soon as instructions are given and user has verified understanding.

Ending point: User types the name of the third data abnormality into the appropriate text entry box.

Possible errors:

- User fails to find the “Lab Data” screen
- User fails to identify which data are abnormal.
- User fails to use the pull-down menu to mark the data as high, normal or low.
- User fails to enter the names of the data abnormalities in the appropriate text boxes.

Benchmark Task #6

Task Statement: For Case 11, create a problem list that explains as well as possible, all abnormal data. Tell the observer when you are done.

Objective: This benchmark task is to verify that the interface makes clear how to generate a problem list, and how to imply causation through the relationship of certain mechanisms/data abnormalities to each other.

Steps involved:

- Open the “View” menu (unless using the toolbar).
- Select the “my problem list” menu item from the menu, or the “my problem list” icon on the toolbar.
- In the “My Problem List Box”, type “Splenic Contraction” and label it as a mechanism by selecting the stylized “M” from the small pull-down menu that will appear to the right of the words just typed.
- Click on the word “Polycythemia” in the “The Data Abnormalities I’ve Found” box and drop it under and to the right of the words “Splenic Contraction” in the “My Problem List” box.

Starting point: Case 11 “Case Summary” screen. Time measurement starts as soon as instructions are given and user has verified understanding.

Ending point: User indicates verbally that the task is done.

Possible errors:

- User fails to recognize that the “My Problem List” screen is the appropriate one for performing this task.
- User fails to understand that the name of the mechanism must be typed.
- User fails to understand that items can be dragged and dropped from one box to another.
- User fails to drag the name of the data abnormality to a position indicating that it is subordinate to the mechanism.

Benchmark Task #7

Task Statement: Create a note in your problem list indicating that polycythemia was relative due to splenic contraction–epinphrine response.

Objective: This benchmark task verifies that the mechanism for creating notes is clear and usable.

Steps involved:

- Click on the arrow to the left of the word “polycythemia” in the “My Problem List” box, causing the arrow to point down.
- Where the flashing text cursor appears, indented and below the word “polycythemia”, type the note.

Starting point: My Problem List screen. Time measurement starts from the moment instructions provided are read by the user.

Ending point: Timing ends when user begins to type note (to allow for slow typists, who understand how to perform the task at hand.)

Possible errors:

- User does not click the arrow down beneath the label “polycythemia.”
- User fails to recognize where to type.

Benchmark Task #8

Task Statement: Report a diagnosis for the patient in Case 11. Once you know the system has recorded your diagnosis, tell the observer.

Objective: This benchmark is designed to show if the interface clearly indicates how to view and compare the problem lists.

Steps involved:

- Open the “View” menu (unless using the toolbar).
- Select the “expert problem list” menu item from the menu, or the “my problem list” icon on the toolbar.
- Compare the expert problem list with “my problem list.”
-

Starting point: Case 11 “my problem list” screen. Time measurement starts as soon as instructions are given and user has verified understanding.

Ending point: User indicates verbally that the task is done.

Possible errors:

- User fails to recognize that choosing the “View Problem List” will allow for comparing it with “my problem list.”

Benchmark Task #9

Task Statement: Report to the system how closely “My Problem List” matches the “Expert Problem List”.

Objective: This benchmark task checks the usability of the evaluate function

Steps involved:

- Click the evaluate button (in the expert problem list window).
- Using the drop-down menu, select the description (and accompanying number) that best describes the relationship between the two problem lists.
- Click on the “submit evaluation” button.

Starting point: Expert Problem List screen.

Ending point: Expert Problem List Screen. Task ends when user clicks on the “submit evaluation” button.

Possible errors:

- User fails to select the “evaluate” button.
- User fails to employ the pull-down menu to select the appropriate rating.
- User fails to click the “submit evaluation” button.

Benchmark Task #10

Task Statement: Without manually searching through “My Problem List” see if the word “anemia” appears in “My Problem List.” If so, indicate to the observer that it does.

Objective: This benchmark is designed to evaluate the usability of the “find” interface.

Steps involved:

- Open the “Case Menu” and select the “Find” menu item, or
- Select the “find” menu item from the toolbar
- Enter the word “anemia” in the text box following the word “find”
- Double-click on the word “anemia” which appears in the “found” box.
- Show the observer where the word “anemia” appears.

Starting point: Any screen after Case 11 has been opened.

Ending point: Timing stops when the user indicates to the observer where the word “anemia” appears.

Possible errors:

- User fails to select the Case menu item
- User selects the wrong option in the case menu
- User does not click the “okay” button to close the case or quit the program.

Benchmark Task #11

Task Statement: Close Case 11 and Exit the PLG.

Objective: This benchmark checks the ease with which the user can close a case and exit the program.

Steps involved:

- Open the Case Menu
- Select the Close menu item
- Click the “okay” button
- Open the Case Menu
- Select the Exit menu item
- Click the “okay” button
- (Note, the first 3 steps can be bypassed, as choosing exit prompts to close the case, and then exits the program.

Starting point: Any screen after Case 11 has been opened.

Ending point: Regular system desktop. Task ends when the program exits.

Possible errors:

- User fails to select the Case menu item
- User selects the wrong option in the case menu
- User does not click the “okay” button to close the case or quit the program.

APPENDIX C -- INFORMED CONSENT FORM

Informed Consent for Participant of Investigative Project

Virginia Polytechnic Institute and State University

Title of Project: On-line Veterinary Medicine Clinical Problem Classifier

Principal Investigator: Jared A. Danielson

I. THE PURPOSE OF THIS RESEARCH/PROJECT

You are invited to participate in the evaluation of the Veterinary Medicine Clinical Problem Classifier. The problem classifier will allow students of Veterinary Medicine, as well as practitioners to practice their skills in classifying problems and arriving at diagnoses given laboratory data. The purpose of this study is to evaluate and improve the problem classifier, making it a valuable and useful learning tool for students and practitioners.

II. PROCEDURES

You will be asked to participate in a one-hour interview/observation session. During this time, you will be asked to participate in a variety of problem classification tasks, either on paper, or using the on-line problem classifier. These tasks will be familiar to you. They involve classifying and organizing lab data in such a way as to be able to arrive at a diagnosis of the problem (if any) suffered by the animal described in each case. Your performance will not be evaluated, nor will anything you do affect how you are assessed academically or professionally. Your participation will help the evaluators 1) understand the process of problem classification and 2) evaluate the effectiveness of the problem classifier software. All information that you help us attain will remain confidential. Your actions will be noted and you will be asked to describe verbally your learning process. You may be asked questions during and after the evaluation, to help us understand what you are doing and thinking.

The session will last about 60 minutes. There are no risks to you. The tasks are not very tiring, but you are welcome to take rest breaks as needed. If you prefer, the session may be divided into two shorter sessions.

III. RISKS

There are no known risks to the subjects of this study.

IV. BENEFITS OF THIS PROJECT

Your participation in this project will provide information that may be used to create and/or improve the problem classifier. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed. If you wish to receive such a synopsis, please leave a self-addressed envelope with the experimenter and a copy of the results will be sent to you.

You are requested to refrain from discussing the evaluation with other people who might be in the candidate pool from which other participants might be drawn.

V. EXTENT OF ANONYMITY AND CONFIDENTIALITY

The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a subject number and/or pseudonym will identify you during analyses and any written reports of the research.

The experiment may be videotaped. If it is taped, the tapes will be stored securely, viewed only by individuals working on the project, and erased after the project is completed. If the experimenters wish to use a portion of your videotape for any other purpose, they will get your written permission before using it. Your signature on this form does not give them permission to show your videotape to individuals not working on the project.

VI. COMPENSATION

Your participation is voluntary and unpaid.

VII. FREEDOM TO WITHDRAW

You are free to withdraw from this study at any time for any reason.

VIII. APPROVAL OF RESEARCH

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Teaching and Learning.

IX. SUBJECTS RESPONSIBILITIES AND PERMISSION

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

Signature

Date

Name (please print)

Contact: phone or address or

E-mail address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigator: Jared A. Danielson

Phone: (540) 231-5587
Graduate Student, Dept. of Teaching and Learning
E-mail: jadanield@vt.edu

Instructor: Dr. John Burton

Phone: (540) 231-5587
Professor, Department of Teaching and Learning
E-mail: jburton@vt.edu
Phone: (540) 231-9359

Review Board: Chair, IRB

Research Division
306 Burruss Hall
Campus 0244
Virginia Tech

xc: the participant, Dr. Burton

HRH: Revised 10/1/97

APPENDIX D: DIRECTIONS SHEET

Instructions for Evaluation of the On-line Problem List Generator (PLG) System

Thanks so much for your participation in this evaluation. We appreciate your willingness to help us evaluate the On-line Problem List Generator (PLG), which we hope will eventually help many Vet Students more easily understand how to classify clinical laboratory data.

The final version of the PLG that you will be using will be accessible from any computer platform using any java-enabled web browser. However, you are being asked to help us evaluate the PLG in one of several forms, including a paper prototype or a variety of electronic versions. Using the PLG entails typing occasionally, and clicking on and/or dragging objects with a mouse. All these activities are probably familiar to you if you have used a computer in the Windows or Macintosh environments.

Using the PLG is much like creating a problem list on paper. It allows the user to classify lab data, name it, and organize it hierarchically, implying causal relationships.

During this session, we'll ask you to create a completed problem list using the system. In order to best evaluate this system, we will be taking notes and observing you as you do this.

Because we want to know whether or not this system is clear and easy to use, we are interested in what you are thinking as you use it. Please feel free to stop and ask questions or make suggestions as you move through the problem list generator. The observer might occasionally ask you questions about why you did certain things you did. All of this data will help us as we improve this system.

Remember, it's impossible to make a mistake! You're helping us test the system; we're not testing you. Please be honest about your impressions -- both good and bad. We want to know what's working well, but we also want to know what to fix.

When you're done with your problem list we will ask you to complete a short questionnaire designed to get your feelings about your experience with the PLG. Again, we thank you for complete openness in your responses. The more honest you are with us, the better the system can be. If any of the questions on the questionnaire aren't clear to you, please feel free to ask the evaluator to clarify them.

The session will last an hour. Once again, thank you for taking the time to help us improve our On-line Problem List Generator.

APPENDIX E -- 1ST SMALL GROUP QUESTIONNAIRE

Usability Questionnaire (Accompanying Usability Specifications):

Please circle the numbers that most appropriately reflect your impressions about using this computer system.

Does Not Apply = NA

- | | | |
|---|---|---------------------|
| 1. Operations relate to tasks | distantly
0 1 2 3 4 5 6 7 8 9 10 | closely
NA |
| 2. Display layouts make tasks simpler | never
0 1 2 3 4 5 6 7 8 9 10 | always
NA |
| 3. Sequence of displays | confusing
0 1 2 3 4 5 6 7 8 9 10 | clear
NA |
| 4. Learning the Operation | difficult
0 1 2 3 4 5 6 7 8 9 10 | easy
NA |
| 5. Human memory limitations | overwhelmed
0 1 2 3 4 5 6 7 8 9 10 | are respected
NA |
| 6. Overall navigation (Getting around in the program) | irritating
0 1 2 3 4 5 6 7 8 9 10 | easy
NA |
| 7(a). Overall reactions | terrible
0 1 2 3 4 5 6 7 8 9 10 | wonderful
NA |
| (7b) | frustrating
0 1 2 3 4 5 6 7 8 9 10 | satisfying
NA |
| (7c) | uninteresting
0 1 2 3 4 5 6 7 8 9 10 | interesting
NA |
| (7d) | dull
0 1 2 3 4 5 6 7 8 9 10 | stimulating
NA |
| (7e) | difficult
0 1 2 3 4 5 6 7 8 9 10 | easy
NA |

**APPENDIX F: SECOND SMALL GROUP AND FIELD TRIAL
QUESTIONNAIRE**

Usability Questionnaire (Accompanying Usability Specifications):

Please circle the numbers that most appropriately reflect your impressions about using this computer system.

Does Not Apply = NA

- | | | | | |
|---|--------------|------------------------|-------------|------------|
| 1. Using the Problem List Generator made me feel. . . | overwhelmed | 0 1 2 3 4 5 6 7 8 9 10 | comfortable | NA |
| 2. Display layouts make tasks simpler | never | 0 1 2 3 4 5 6 7 8 9 10 | always | NA |
| 3. Sequence of displays | confusing | 0 1 2 3 4 5 6 7 8 9 10 | clear | NA |
| 4. Learning How to Use It | difficult | 0 1 2 3 4 5 6 7 8 9 10 | easy | NA |
| 5. Compared to how I do problem lists now, using the Problem List Generator to do cases is. . . | harder | 0 1 2 3 4 5 6 7 8 9 10 | easier | NA |
| 6. Using the Problem List Generator. . . | | | | |
| 6 (a) Made me account for more lab data than I normally do | Less | 0 1 2 3 4 5 6 7 8 9 10 | Same | More
NA |
| 6 (b) Made my problem list more precise | Less | 0 1 2 3 4 5 6 7 8 9 10 | Same | More
NA |
| 6 (c) Made me remember too much | Yes/Too much | 0 1 2 3 4 5 6 7 8 9 10 | No | NA |
| 7. I could get used to using the Problem List Generator to do all my cases | No Way | 0 1 2 3 4 5 6 7 8 9 10 | Definitely | NA |
| 8. Overall navigation (Getting around in the program) | irritating | 0 1 2 3 4 5 6 7 8 9 10 | easy | NA |

9. Overall reactions	terrible		wonderful	
		0 1 2 3 4 5 6 7 8 9 10		NA
(9b)	frustrating		satisfying	
		0 1 2 3 4 5 6 7 8 9 10		NA
(9c)	uninteresting		interesting	
		0 1 2 3 4 5 6 7 8 9 10		NA
(9d)	dull		stimulating	
		0 1 2 3 4 5 6 7 8 9 10		NA
(9e)	difficult		easy	
		0 1 2 3 4 5 6 7 8 9 10		NA
10. I was frustrated by technical problems	No		Yes	
		0 1 2 3 4 5 6 7 8 9 10		NA
11. Doing my problem list homework using the problem list generator will be easier than the way I do it now.	definitely not		absolutely	
		0 1 2 3 4 5 6 7 8 9 10		NA
12. I like being able to do my problem lists on a computer.	definitely not		absolutely	
		0 1 2 3 4 5 6 7 8 9 10		NA
13. I like being able to do my problem lists on the web.	definitely not		absolutely	
		0 1 2 3 4 5 6 7 8 9 10		NA
14. I like having my problem lists turned in automatically as soon as I finish them.	definitely not		absolutely	
		0 1 2 3 4 5 6 7 8 9 10		NA
15. Using the problem list generator makes it easier for me to organize my thoughts about a case.	definitely not		absolutely	
		0 1 2 3 4 5 6 7 8 9 10		NA
16. Using the problem list generator will make understanding clinical pathology. . .	harder		easier	
		0 1 2 3 4 5 6 7 8 9 10		NA

17. What do you like most about using the problem list generator?

18. What do you like least about using the problem list generator?

19. What would you change about the problem list generator if you could?

20. Any additional comments you'd like to make about the problem list generator:

For questions above which you ranked particularly low, please indicate why here:

WORKS CITED -- APPENDICES A THROUGH F

- Dick, W., & Carey, L. (1996). The systematic design of instruction. (4th ed.). New York, NY: HarperCollins College Publishers.
- Fitz-Gibbon, C. T., & Morris, L. L. (1978). How to design a program evaluation. Beverly Hills, CA: Sage Publications.
- Hix, D., & Hartson, H. R. (1993). Developing user interfaces: Ensuring usability through product and process. New York: John Wiley & Sons, Inc.
- Hix, D., & Hartson, H. R. (1997). Usability engineering course packet: CS 5714: Fall 1997. Blacksburg, VA: Virginia Tech.
- Mandin, H., Jones, A., Woloschuk, W., & Harasym, P. (1997). Helping students learn to think like experts when solving clinical problems. Acad Med, 72(3), 173-179.
- Tessmer, M., & Richey, R. C. (1997). The role of context in learning and instructional design. Educational Technology Research and Development, 45(2), 85-115.