2013-12-13

Novice and Expert Differences and Educational Interventions to Improve Veterinary Pathology Visual Diagnostic Reasoning Measured by Eye-tracking Technology

Warren, Amy Louise

UNIVERSITY OF CALGARY

Novice and Expert Differences and Educational Interventions to Improve Veterinary Pathology Visual Diagnostic Reasoning Measured by Eye-tracking Technology

by

Amy Louise Warren

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MEDICAL SCIENCES

CALGARY, ALBERTA

DECEMBER 2013

Abstract

Purpose: There were two objectives, to 1) to use eye-tracking to establish baseline quantitative and qualitative differences between novice and expert veterinary pathologists and explore dual process theory of clinical reasoning, and 2) determine if the introduction of two educational interventions, the active use of key diagnostic features and image repetition, improved novice visual diagnostic reasoning skills.

Method: A pre-experimental static group comparison between novice and expert veterinary pathologists was used. Participants were shown 10 veterinary cytology images and asked to formulate a diagnosis while wearing eye-tracking equipment (10 slides) and while concurrently thinking aloud (5 slides). A quasi-experimental, pre-test and post-test comparison group design was used to compare the two teaching interventions to a comparison group using eye-tracking as an assessment method. The time to diagnosis and percentage time spent viewing an area of diagnostic interest (AOI) were compared using independent *t*-tests (novice and expert) and paired *t*-tests (time) and analysis of covariance (ANCOVA) (between groups) was used for the educational interventions. Diagnostic accuracy as a dichotomous variable was compared using chi-square tables.

Results: Compared to novice, experts demonstrated significantly higher diagnostic accuracy ($p < 0.017$), shorter time to diagnosis ($p < 0.017$) and a higher percentage of time spent viewing AOIs ($p < 0.017$). Experts elicited more key-diagnostic features in the think-aloud protocol and had more efficient patterns of eye-movement. Students in the extended visual reasoning teaching intervention: active learning, image repetition behaved most like experts with no significant difference to experts for diagnostic accuracy, percentage time spent in the AOIs and a significantly faster time to diagnosis than experts ($p < 0.017$).

Discussion: I suggest that experts' fast time to diagnosis, efficient eye-movement patterns, and preference for viewing AOIs supports system 1 (pattern-recognition) reasoning and script-inductive knowledge structures with system 2 (analytic) reasoning to verify their diagnosis. Our results from the educational interventions suggest a greater level of improvement in the eye-tracking of students that were taught key-diagnostic features in an active learning forum and were shown multiple case examples.

ACKNOWLEDGEMENT

I think all higher degrees are accomplished with a "village" of supporters. To my supervisor, Tyrone Donnon- thank you for your wonderful mentorship and support and for entertaining my crazy ideas. To my supervisory committee: Dr. Tanya Beran, Dr. Kent Hecker and Dr. Claudio Violato – thank you for your insight and wisdom. Particular thanks to Dr. Kent Hecker for his encouragement in this project. I completed this PhD while working, raising my now 5 year-old daughter and being pregnant with and having my now 1 year-old son. I could not have even embarked on this without the undying support and championing of my wonderful, other half of the sky, Robin Yates- I love you. All the nights when I came home ready to give in, you were there to tell me I could do it, picked up our kids when I had to work late and gave me both emotional and intellectual support, ideas and encouragement. You are my rock. To my wonderful babies, Charlotte Lilly and Elliot James, I love you both and know the sacrifices you've both made when I have had to share my time with you with a PhD- Charlie, you must be one of the few five year olds who pretend to be a "PhD student" in your make-believe games. My mum, Elizabeth Warren, is a professor in mathematics education. She taught me the love of educating and the joy of giving back to a profession. Thank you for your support, love, encouragement and ideas through this process. To Jason Abboud, my technician, thank you for your enthusiasm. It has been wonderful working with some one as dedicated and hard working as you.

DEDICATION

This thesis is dedicated to my husband Robin, my daughter Charlotte, and my son Elliot. I love you guys.

Table of Contents

LIST OF TABLES

Table Number and Title **Page No.**

LIST OF FIGURES

Figure Number and Title **Page No.**

LIST OF SYMBOLS, ABBREVIATIONS AND NOMECLATURE

Symbol/ Abbreviation	**Definition**
ACVP	American College of Veterinary Pathologists
ANCOVA	Analysis of Co-variance
AOI(s)	Area(s) of Interest
ASL	Applied Science Laboratories
BVSc	Bachelor of Veterinary Science
CH	Cutaneous Histiocytoma
χ	Pearson Chi-square Value
CL	Cutaneous Lymphoma
d	Cohen's *d* effect size
df	Degrees of Freedom
Dipl ACVP	Diplomate of the American College of Veterinary Pathologists
DTU	Data Transmit Unit
DVM	Doctor of Veterinary Medicine
ET	Eye-tracking
HD	High Definition
LCD	Liquid Crystal Display
MA	Massachusetts
M	Mean
MCT	Mast Cell Tumor
Max.	Maximum

Min.	Minimum
n	Sample Size
p	Two-tailed Significance
%	Percentage
PhD	Doctor of Philosophy
PCT	Extramedullary plasmacytoma (plasma cell tumor)
s	Seconds
SD	Standard Deviation
SE	Standard Error
t	*t*-value
TA	Think-aloud
TVT	Transmissible Venereal Tumor
USA	United States of America

Chapter One: **Introduction**

1.1 Overview

Visual diagnostic reasoning is an essential skill clinicians develop during their veterinary training. Reasoning capability is based on the integration and effective application of thinking and learning skills to generate knowledge within familiar and unfamiliar clinical experiences.[3] Cytology, a sub-specialty of veterinary clinical pathology, relies on the integration of visual cues from cytologic specimens to form a diagnosis. In certain veterinary and medical specialties, including pathology, diagnostic reasoning is almost exclusively reliant on visual cues, "visual diagnostic reasoning".

Previously, due to the physical intimacy of the pathologist and microscope, diagnostic pathology reasoning processes were assessed based on think-aloud protocols and identified several discrete steps in the acquisition of competence: development of adequate search strategies, rapid and accurate recognition of anatomic location, acquisition of visual data interpretation skills, and transitory reliance on explicit feature identification.[4] Think-aloud protocols, while informative to some aspects of clinical reasoning, are hampered by their poor ability to access non-analytic parts of the reasoning process (particularly pertinent to visual reasoning) and have been shown to hamper the speed of visual diagnosis, thus altering the reasoning process. Recent studies in human radiology and dermatology have utilized eye (gaze) tracking technology to assess visual diagnostic reasoning skills.[5] With the development of virtual microscopy utilizing instruments such as the Aperioscope, eye tracking technology is now available to pathologists to study visual diagnostic reasoning.[6, 7] To date, minimal research has been conducted to assess visual diagnostic reasoning in veterinary medicine or veterinary

pathology, and little research has been conducted in human pathology visual diagnostic reasoning. Understanding the visual diagnostic reasoning processes employed by experts (veterinary pathologists) is informative to teaching visual diagnostic reasoning of novices (veterinary students).

Lessons from the medical literature on visual diagnostic reasoning (in radiology and dermatology) suggest that differences in novices and experts are related to the total time in which images are viewed, selection of fixation points and time spent on fixation points. Using these three outputs, I aimed to investigate visual diagnostic reasoning in veterinary pathology using eye-tracking technology.

1.2 Purpose of the study

Improving student's diagnostic reasoning abilities has enormous benefits to high-stake decision fields of veterinary and human medicine. The purpose of this study was three-fold. First, I aimed to validate the use of eye-tracking technology in studying veterinary pathology diagnostic reasoning skills using cytology specimen images. Second, I hypothesized that expert participants would take less time to scan a slide, be more accurate in identifying diagnostically useful points of interest, spend more time visualizing points that relate to their diagnostically useful points of interest and have higher diagnostic accuracy. Third, I aimed to use eye tracking to assess two educational interventions: the use of explicit visual features and image repetition.

I hypothesized that visual diagnostic reasoning in experts and novices has components of both System 1 and System 2 cognitive processes. This is, in part, supported in the visual diagnostic reasoning literature in radiology. Further, I hypothesized that experts develop sophisticated visual "illness scripts", where visual images attained with experience are the trigger

of a knowledge structure with associated slots of disease probability, pathophysiology, anatomy and prognostic outcomes. In Objective 1bii, I aimed to investigate baseline differences between novices and experts using eye tracking. By using the eye-tracking data, I aimed to demonstrate that experts had more efficient and targeted viewing of an image, akin to a System 1 thinking process.

Based on educational principles of clinical reasoning, veterinary pathology diagnostic reasoning may also benefit from educational interventions that include repeated and directed exposure of students to cases that are active, integrate prior biomedical knowledge and provide immediate instructor feedback.[8-13] The teaching of knowledge in "schemes" has been suggested to enhance memory organization and diagnostic success.[14] This is supported in radiology visual reasoning where expert radiologists construct more elaborate and flexible schemata only a short time after exposure to the film in comparison to novices.[15] Based on this, I aimed to investigate whether the use of explicit diagnostic features improved novice visual reasoning skills when presented with static cytology images. Secondly, I aimed to investigate whether directed image repetition improved visual diagnostic reasoning skills (Objective 2).

As investigating diagnostic reasoning in veterinary pathology is a new field, the preliminary studies outlined in this thesis will lay the ground work for future studies into visual diagnostic reasoning processes and teaching in veterinary programs.

1.3 Research Questions

Is there a quantifiable difference in visual diagnostic reasoning performance between expert and novice veterinary pathologists using eye-tracking technology? Using concurrent think-aloud protocols with eye tracking, are there differences between novices and experts use of

System 1 and System 2 reasoning processes? Does the introduction of two educational interventions facilitate visual diagnostic reasoning development in novice (DVM student) veterinary pathologists?

1.4 Objectives of the Study

1. To quantitate visual diagnostic reasoning performance in expert and novice veterinary pathologists
    a. To investigate the use of eye-tracking methodology in measuring visual diagnostic reasoning skills by comparing it to established think-aloud protocols and measures of diagnostic success.
    b. Novice and Expert studies:
        i. To use eye tracking technology as a measurement tool for visual diagnostic image evaluation and to establish baseline differences in eye tracking ability in novice vs. expert pathologists.
        ii. To investigate dual process theory as a potential continuum between novice and experts based on think-aloud recordings and eye-tracking data.
2. Visual reasoning educational interventions.
    a. To determine if the use of explicit features in image interpretation improves novice pathologist visual diagnostic reasoning skills.
    b. To determine if the introduction of pathology image repetition improves novice pathologist visual diagnostic reasoning skills.

Chapter Two: **Literature Review**

2.1 General

The central importance of pathology to the understanding and diagnosis of human and veterinary disease is little contested.[16] Pathology, the study of disease mechanisms and manifestations, links normal and abnormal, tissue function and dysfunction and the basic sciences and clinical medicine. It has a fundamental role in the construction of medical and veterinary knowledge and understanding, as well as being a diagnostic specialty.[17] In veterinary medicine, pathology is subdivided into two main disciplines, clinical pathology, encompassing hematology, cytology, clinical chemistry and clinical immunology; and anatomic pathology, focusing on tissue pathology: gross and histopathology. Concurrent with the expanding understanding of molecular biology, molecular pathology, the study of disease at the level of cells, organelles and genes, has had an increasing emphasis in veterinary pathology education.[18]

Cytology is a subspecialty of clinical pathology and has advantages of being minimally traumatic to the patient (done in the clinic with no sedation required), fast and relatively inexpensive.[19] For the most part, a specialist veterinary clinical pathologist performs detailed diagnostic analysis of cytological preparations. However, due to the accessibility of fine needle aspirates to clinicians and the ease of in-practice staining, general practitioners in clinics will diagnose subsets of tumors and disease processes cytologically. Although a specialist skill, cytologic diagnosis is also taught at a rudimentary level to Doctor Veterinary Medicine (DVM) students.

Cytologic samples are attained by fine needle aspiration. Fine needle aspirates are fixed and stained (typically with Wrights Giemsa or Diff-Quik stain) and examined microscopically. Cytology allows pathologists to detect abnormalities in tissues at the cellular level. Visual

disease manifestations include changes in expected cell populations and proportions, architectural alterations, cytonuclear abnormalities and molecular changes visualized using molecular markers or stains (i.e. immunohistochemistry, fluorescent *in situ* hybridization, specialized histochemical staining).

Interpreting cytology slides requires a complex integration of visual cues (principally pattern recognition), detailed microanatomy and physiology knowledge, and an intimate understanding of pathologic processes.[20] The diagnostic reasoning underpinning decision-making in cytology integrates complex visual data and intellectual processes in a "visual reasoning" process.[4] A pathologist gathers all the data utilized in diagnostic reasoning simply by changing their point of gaze. In this, the eye is the "window" to the reasoning process and much information about how diagnostic reasoning occurs can be gathered by tracking eye-movement.

2.2 Clinical and Diagnostic Reasoning

2.2.1 General Background

Clinical (and diagnostic) reasoning, a critical skill clinicians develop through their education and career, relies on the integration and application of thinking and learning skills to make decisions in familiar and unfamiliar clinical experiences.[21,3] It relies on both theoretical and experiential knowledge. Experiential knowledge is built as new experiences force the revision of past knowledge and skill development.

Several theories about the cognitive process of diagnostic reasoning have been proposed over the years: hypothetico-deductive, knowledge structures, script theory, pattern recognition and dual process theory. Despite multiple attempts to ascertain a single reasoning process, it is more probable that experts' reasoning ability is the result of an extensive and multidimensional

knowledge base and that multiple reasoning processes are employed, based on the context of the case.[12, 13]

2.2.2 Hypotheticodeductive Model

Hypothetico-deductive reasoning, proposed in the 1970s and 1980s,[22] is similar to the reasoning used in scientific investigation where knowledge grows through a succession of tested hypotheses for which experimental evidence is sought to prove or disprove.[23, 24] It was thought that clinical reasoning utilized this generic problem solving strategy. Backward reasoning underpins hypothetico-deductive reasoning and is characterized as reasoning from the diagnosis (hypothesis) back to the data.[25] Forward reasoning, used more frequently by experts, starts with observations and data to generate a diagnostic hypothesis.[25] Hypothetico-deductive, forward and backward reasoning are all analytical reasoning processes.

Two studies cast doubt on the generalizability of hypothetico-deductive reasoning. Firstly, although similar methods of problem solving were employed, experts' generation of hypotheses, while not quantitatively different from novices, were qualitatively better.[26] Secondly, success in one problem correlated poorly with success in a second problem in a different area. This suggested there was content specificity to problem solving.[22] While hypothetico-deductive reasoning likely contributes to some clinical reasoning in experts, this model failed to explain how experts generated the initial hypothesis and how knowledge and knowledge structures influenced the process.

2.2.3 Knowledge Structures

Emerging from the findings that clinical reasoning did not use a single generalizable problem solving method were studies that examined the context specificity of reasoning, in particular the knowledge differences between novices and experts. Using propositional analysis, Patel and Groen[27] (1986) demonstrated that diagnostic success and expertise were associated with forward reasoning and that backward reasoning tended to be a weaker reasoning modality. Clinicians using forward reasoning relied on causal rules derived from their underlying knowledge base. It is expected that the more knowledge experts have, the more information they will retrospectively recall. Several studies in the 1980s, however, cast doubt on this assumption and an "intermediate effect" was demonstrated where intermediates, when asked to recall information from a case they had read, were able to recall more information from the case than novices and experts.[27] Norman et al.[28] (1989), however, later showed that when experts and novices were challenged to *interpret* laboratory data, forcing abstract mental representation, experts were able to recall more information on the case than novices. Norman suggested that earlier experiments that demonstrated an "intermediate effect" reflected different processing strategies; novices relying more on analytical processing, experts on pattern-recognition.[28]

Boshuizen and Schmidt[29] (1992) also demonstrated through think-aloud and post-hoc explanation, that experts had more in-depth biomedical knowledge than novices and intermediates, and suggested that knowledge expertise developed in three stages- acquisition, practical experience and encapsulation (integrating theoretical and experiential knowledge). Experts made more inferences from fewer data points in the text as experts were better able to distinguish diagnostically useful data from the case.[30] When novices, intermediates and experts were primed with biomedical knowledge prior to reading a case and recalling it an intermediate

effect was still observable, evidence Schmidt and Boshuizen[30] (1993) concluded supported the encapsulation theory. The encapsulation theory is supported in later studies by Rikers et al.[31] (2005) and de Bruin et al.[32] (2005) using a lexical decision study and structural equation modeling, respectively.

2.2.4 Script Theory

In 1990, Schmidt et al.[33] proposed a theory of clinical reasoning that was based on the premise that expertise was not reliant on in-depth knowledge or reasoning skills, but was based on the *knowledge structures* that a clinician possesses. Building on a cognitive science theory of "schemes", they proposed that similar cognitive structures, "illness scripts" were possessed by experienced clinicians.[26, 33-35] Illness scripts are knowledge structures constructed during repeated experiences that organize medical knowledge into readily mobilized schemes that are goal-directed and increase the efficiency of task performance.[33] Illness scripts are activated when a practitioner first perceives a patient's symptoms and interprets characteristics and features of a case to infer a differential diagnosis.[35] Each generated hypothesis (i.e., differential diagnosis) is an activated illness script. Illness scripts have hierarchical knowledge structures "slots" containing relevant medical information pertaining to the clinical signs, including pre-stored knowledge about different predisposing conditions, pathophysiology and diagnostic testing, and acceptable or not acceptable values and default values.[34-36] Once a selection of scripts (or script) are identified as working clinical hypotheses, data collected is used to ascertain if the clinical signs fall within the pre-stored acceptable values for the script.[35] If unacceptable values are identified for a clinical sign or test, the script is rejected. Clinical signs and test results

within a script are also weighted in that some clinical signs are more indicative of a diagnosis than others.[35]

Based on this, Schmidt et al.[33] (1990) proposed there were four stages to developing illness scripts in clinical reasoning: 1) elaborated causal networks 2) abridged networks 3) illness scripts and 4) instance scripts. Elaborated causal networks link concepts based on cause and consequence, including linking biomedical aberrations to presenting clinical signs or pathophysiologic consequences.[37] As students are exposed to extensive and repeated application of causal networks, abridging of the networks occur. From this high-level, simplified causal models develop amalgamating signs and symptoms into a diagnosis. As students' experiences with cases grow, they develop illness scripts and later, as experienced clinicians, instance scripts. It is thought that the repertoire of illness scripts that clinicians acquire with experience allow problem solving by recognizing the presenting problem as being similar to ones previously solved (i.e., pattern recognition).[14]

The theory of illness scripts and their development explains some of the data inconsistencies in previous studies. Firstly, the biomedical knowledge structures and experiences that a clinician encapsulates in an illness script are specific to a domain, explaining the "content specificity" of problem solving ability in experts. Secondly, experts remember fewer data features in a case description as they rapidly recognize a problem as being similar to a previous illness script. In addition, the intermediate effect can be explained by expert knowledge being compiled. This allows experts to focus only on critical aspects of the problem, and, thus, they remember fewer data points and make fewer pathophysiologic inferences.[30] Intermediates, on the other hand, are at the elaborated knowledge stage, thus activate more pathophysiological networks when presented with a problem and recall more data.[37]

2.2.4.1 Role of Biomedical Science Knowledge

Considering the important role of biomedical science knowledge in script construction, there were several studies in the 1980s and 90s that explored the role of biomedical science knowledge in clinical reasoning. Most studies used written cases as a stimulus for participants to discuss their reasoning or provided a basic science test and clinical cases to assess participants' integration of basic science concepts in clinical reasoning. Think-aloud protocols and propositional analysis allowed researchers to identify basic science concepts. Using this experimental format, Patel and Groen[27] (1986) and Boshuizen and Schmidt[29] (1992), demonstrated that experts used less biomedical knowledge in their clinical reasoning, leading some to argue that biomedical knowledge was redundant and separate from clinical knowledge. Further studies by Norman et al. [17] (1994) and Woods et al. [31] (2007), using similar experimental formats but altering the tasks provided (i.e., problem difficulty and differential prior training), demonstrated that experts actually have more in depth biomedical knowledge and that their knowledge is encapsulated, and hence not "visible" in think-aloud protocols. Norman et al.[25] (1994) also demonstrated that experts made extensive use of biomedical concepts (mostly in well-organized chunks) when solving complex problems, providing further support for the knowledge encapsulation theory. Work by Rikers et al. [23] (2005) and de Bruin et al. [24] (2005) also support the premise that biomedical knowledge is encapsulated into clinical knowledge by experts.

2.2.5 Pattern Recognition

Emerging simultaneously with work on knowledge structures was the recognition that while analytical styles of reasoning explained some of the clinical reasoning process, particularly in novices, experts used a different form of reasoning that was intuitive and pattern based.[36, 38] Analytical reasoning analyzes a case on a feature-by-feature basis whereas in non-analytical processing is holistic (i.e., the global impression of the case is compared to prior experienced exemplars).[39] In pattern recognition models of reasoning, case presentations resemble previously seen disease examples and trigger disease categorization and automatic knowledge retrieval from well-structured knowledge patterns.[35] Papa et al. [38] (1990) used a prototype-driven medical decision model to assess diagnostic accuracy due to pattern recognition and demonstrated that diagnostic accuracy was determined by the ability to discriminate between patterns rather than the ability to match patterns. Similar theories of clinical reasoning including scripts[37] use a pattern-based approach to reason in an intuitive manner. The two seemingly dissonant groups of reasoning, those reliant on analytic skills and the other pattern based, can be explained by a psychological theory of decision making: dual processing theory.

2.2.6 Dual Processing Theory

A prominent emerging theory of human judgment is dual process theory where two systems of cognition are proposed to exist: System 1) heuristic and intuitive, and System 2) systematic, deliberate and analytical.[40, 41] Numerous attributes have been ascribed to the systems, reflecting levels of consciousness, evolution, functional characteristics and individual differences (Table 2.1).[1] System 1 cognition is generally fast, high capacity, unconscious (first impression), perceptive, contextualized, associative and highly reliant on pattern recognition. It

is also prone to error and due to its context-bound nature, prone to subconscious influences of emotion and prior experiences.[42] In contrast, System 2 cognition is slow, deliberate, analytical and conscious, and is less prone to error.[42] As System 2 is relatively slow and effortful, it is susceptible to interference associated with time pressure, fatigue, cognitive difficulty of a task, and competing thought processes.[21, 43]

While two delineated systems have been proposed, cognitive theorists suggest that a continuum (also referred to as cognitive continuum theory) rather than a dichotomy exists and that these systems likely act in parallel during the reasoning process.[41, 43-45] While System 1 (intuitive) is a default response, it can be over-ridden or altered by analytical reasoning (System 2).[46] Triggering of analytical reasoning occurs with strong deductive reasoning instructions, dissonant patterns and tends to occur more often in individuals with high cognitive ability.[9, 46] System 2 reasoning is believed to be a gate-keeper, monitoring System 1 reasoning processes.[47]

Table 2.1. Attributes of System 1 and System 2 thinking (adapted from Evans et al., 2008)[1]

Attribute cluster	System 1	System 2
Consciousness	Unconscious	Conscious
	Implicit	Explicit
	Automatic	Controlled
	Low effort	High effort
	Rapid	Slow
	High capacity	Low capacity
	Holistic, perceptual	Analytic, reflective
Evolution	Evolutionally old, shared with animals	Evolutionally recent, uniquely human
Functional characteristics	Associated	Rule-based
	Contextualized	Abstract
	Pragmatic	Logical
	Parallel	Sequential
Individual differences	Independent of intelligence and working memory	Linked to intelligence and limited to working memory capacity

In the context of diagnostic reasoning, System 1 (intuitive) is highly related to experience.[42] System 2 (analytic) tends to be used by novices, relying on hypothetico-deductive, backward reasoning,[3] aborization (algorithmic), and critical and logical thought.[42] It is thought that experts revert to System 2 (analytic) thinking when faced with unknown or ambiguous presentations,[42, 43, 48, 49] relying more heavily on pathophysiological causality and biomedical knowledge.[21] In the acquisition of highly skilled performance, there is a transformation from System 2 (effortful) to System 1 (effortless) activity[43] with repetition, highly relevant to novice-expert learning.[1, 42] System 1 reasoning is thought to be based on visual similarity and to a lesser degree verbal descriptions,[8] pertinent to fields where visual reasoning predominates (e.g., dermatology, radiology and pathology).[21]

It is tempting to believe that experts' preference for System 1 reasoning indicates a "better" reasoning process, however both systems are likely to be engaged actively in the clinical or diagnostic reasoning process.[11, 12, 50] In the diagnostic reasoning process, the initial fast, automatic generation of a hypothesis or intuitive retrieval of an illness script, is premised on a System 1 process.[21] Generation of greater than one script or diagnostic hypothesis, dissonance between illness scripts, or lack of an exemplifying illness script triggers a deeper System 2 cognitive processes.[35] The clinical examination and diagnostic process (i.e., hypothesis verification) aims to reduce the likelihood of different activated illness scripts (e.g., rule-in or rule-out diagnostic hypotheses) through a System 2 process.[35] Marcum et al.[50] (2012) proposes an integrated model of dual process theory and metacognition, where clinical reasoning starts with a System 1 process to formulate a differential diagnosis, uses a System 2 process to assess the differential diagnoses and formulate a diagnosis and metacognition to reinforce or alter the cognitive process. While not adding a great deal to the granularity of our understanding of

clinical reasoning, dual process theory adds coherence to the observations of clinical reasoning made over the years.

2.2.7 Context of Clinical Reasoning

Recently, theories of ecological or situated cognition have been explored in the clinical reasoning processes of medical experts.[51, 52] The central tenet of these theories is that knowledge is situated in the "experience", referring to not only the participants in the experience but also the physical environment. In a clinical situation, it has been shown that the clinical reasoning of experts was influenced by the context of the presenting case.[51] Further, context appeared to affect diagnostic reasoning more than therapeutic reasoning.[52]

Despite multiple attempts to ascertain a single reasoning process used by clinicians, it is more probable that experts' reasoning ability is the result of an extensive and multidimensional knowledge base and that multiple reasoning processes are employed, based on the context of the case.[12, 13] It is likely that experts in clinical reasoning in general employ a combination of system 1 (pattern recognition) and system 2 (analytic) reasoning in parallel during their diagnostic reasoning process. It is also likely that using system 1 reasoning, they activate a subset of illness scripts, activating knowledge structures. Subsequent to this, an analytical process, using knowledge structures, rules in or out diagnostic hypotheses and employs checks and balances in a metacognitive manner to verify the diagnosis. It is also likely that although visual diagnostic reasoning differs from clinical reasoning in the case presentation, similar cognitive processes are occurring.

2.3 Visual Diagnostic Reasoning

In veterinary and medical fields of pathology, radiology and dermatology, the presentation of a case is highly visual and holistic and, therefore, diagnostic reasoning is heavily reliant on perceptive skills.[53] In contrast to other medical fields, case data is not presented as discrete findings, but is a highly configured visual array of anatomical and physiological features, all instantly available to the observer simply by changing the direction of the gaze.[54] In experts, this process occurs very rapidly with little conscious thought and information can be conceptualized and processed at perceptual speed.[54]

2.3.1 Visualization

Visual sensation, perception and cognition are key elements to visual diagnostic reasoning.[20] Visual sensation and perception are inter-linked: sensation detects the simple properties of colour and light, and perception converts these into perceived objects and images (cells, nuclei, anatomic structures) and background. Vision is a piecemeal process where small visual regions are integrated into a coherent representation of the whole.[55] It is thought that during visual examination, a cyclic repetition of the following steps occur: 1) the entire scene is seen mostly through the peripheral vision (at low resolution) and interesting features "pop out", 2) attention to the entire scene is disengaged and the eyes are repositioned to the first region that attracted attention, and 3) fovea is then directed at the region of interest and foveal attention is engaged at higher resolution.[55] Central foveal (i.e., high resolution) vision is within 1 to 5 degrees of the visual axis and, in eye movement studies, is termed as a "fixation" point or area. When visual attention is directed to a new area, the fovea is repositioned in a fast eye movement "saccade".[55] Although other eye movement types are recognized (e.g., smooth tracking of a

moving object and nystagmus), saccades and fixations predominate during the evaluation of a static image. Areas of fixations correspond to a key characteristic or feature associated with the image the subject wants to maintain attention to. Saccades occur when the subject voluntarily changes the focus of attention from one feature to another.[55]

During visualization, sensory, short and long-term memory allows us to compare images to help us understand what we see. Visual cues trigger strong pattern recognition (i.e., System 1) responses.[42] Using think-aloud reasoning protocols, Crowley et al.[4] (2003) identified five stages during the examination of a histology slide: 1) data examination, 2) data exploration and explanation, 3) data interpretation, 4) control processes, and 5) operational processes. When a pathology slide is examined, the normal anatomy, variations in anatomy, and anatomic variations are rapidly identified.

2.3.2 Expertise in Visual Diagnostic Reasoning

Novice-expert studies in visual diagnostic reasoning are most prevalent for the fields of radiology and dermatology. Expertise in radiology is believed to lie in the experience of viewing thousands of radiology patterns and their synthesis into a coherent, organized and searchable cognitive matrix of diagnostic implications and pathologic features.[56] Akin to "illness scripts" visual diagnosticians develop schema linked to the visual presenting pattern. Expertise in radiology is characterized by three major features: 1) recognition of patterns of abnormality, 2) selection of appropriate diagnostic schema to fit the observations, and 3) maintenance of flexibility in rethinking the decision when new data are presented.[56]

Shortly after viewing a film, an expert radiologist selects an appropriate schema that directs much of the subsequent cognitive processes. The expert tests the schemata until a

diagnosis can be reached, then seeks to confirm it through further testing and data collection.[56] Expert radiologists construct more elaborate and flexible schemata only a short time after exposure to the radiographic image, in comparison to novices. Experts are able to consider more features, more accurately than novices.[15] Novice radiologists have more difficulty building schema and applying them as the testing or information they possess is incomplete. In addition, novices seemed less able to modify schema in response to additional or dissonant information, compared to experts who were flexible and innovative in schemata modification.[56] Experts, further, excel in visual recall, compared to novices. They are able to have more refined schemata of recognition and are capable of finer discriminations.[56]

2.4 Measurement of Visual Diagnostic Reasoning

2.4.1 Novice-Expert Studies

The challenge in studying and measuring diagnostic reasoning is the latent and complex nature of the construct. To date, much of the experimental work on diagnostic reasoning was done using novice-expert studies. Exploring the differences between novices and experts gives insight into how reasoning skills are developed and how experienced clinicians think about the processes they used to derive at a correct diagnosis. A number of different experimental approaches have been used in novice-expert studies, most of which are variations on the research design of case presentation, interpretation and explanation.[57] This includes manipulating features of the case as an independent variable including case difficulty,[58] similarity to previously seen cases,[39, 59, 60] case ambiguity,[61] discordant data,[62] and the addition of irrelevant data.[8] Common dependent variables include diagnostic accuracy,[57] time to reach a diagnosis,[61]

number and order of propositions,[63] directionality of reasoning,[63] knowledge networks,[64, 65] and non-analytic or analytic reasoning modalities.[8]

2.4.2 Think-Aloud and Protocol Analysis

2.4.2.1 Think-Aloud Protocol and Free Recall

In think-aloud and free-recall studies, participants are given a case to study, after which they are asked to recall case information and provide a diagnosis.[28, 30] Studies using free-recall assume the recall data reflect the structure of the mental representation generated by the stimulus.[30] Think-aloud and stimulated recall were insufficient to access non-analytical reasoning[8] and were hampered by what was later described as knowledge encapsulation in experts. Thus, using these protocols, it was erroneously thought that experts possessed less biomedical knowledge resulting in an effect where intermediates appeared to outperform their more experienced expert colleagues.[30] Using case recall as an investigative strategy had limitations too. Although it may have been able to demonstrate differences in the quantity of knowledge possessed by experts and novices, it was unable to explore the knowledge structure (a feature much more pertinent to problem solving and diagnostic reasoning).

2.4.2.2 Explanation Protocols

Explanation protocols are a variant of think-aloud protocols where the participants are specifically instructed to explain the pathophysiologic underpinnings of a case. This was first used by Patel and Groen[27] (1986) to isolate reasoning processes (including directionality) and identify knowledge structures. In diagnostic explanation tasks, subjects are first asked to read a case, then to recall anything they can about the case (free-recall), to provide pathophysiological

explanations for the clinical data, and finally to provide a final or differential diagnosis.[57] Using a diagnostic explanation task, Boshuizen and Schmidt[29] (1992) demonstrated that experts rarely used biomedical knowledge in their reasoning process; however, the biomedical knowledge they possessed was global and comprehensive, which emerged later in a theory of knowledge encapsulation. Schmidt and Boshuizen[30] (1993) demonstrated that although the quantity of concepts recalled by experts was less, the concepts recalled encapsulated several pieces of information in text, forming more accurate inferences. They also demonstrated that when time was restricted, experts, due to their encapsulated knowledge, were able to recall more than intermediates.[30] Diagnostic explanation tasks make several assumptions. Firstly, the case is believed to be processed linearly in that initially information is processed through the working memory and later linked to the long-term memory. Secondly, the temporal sequence of responses from the participant reflects the temporal order of the cognitive reasoning process. Thirdly, the solution strategies and inferences made by the participant reflect underlying knowledge structures.[63]

2.4.2.3 Propositional, Sematic Qualifiers and Semantic Network Analysis

In the process of deriving meaning from the participants' responses, data analysis protocols such as propositional analysis, semantic qualifiers and semantic networks analyze the verbalizations from explanation protocols. Propositional analysis draws conclusions about the knowledge used during problem solving, semantic qualifiers delineate the knowledge structure, and semantic networks provide information on the reasoning process.

Propositional analysis of the diagnostic explanation protocols codes verbalizations into ideas (e.g., surface representations of the text) and propositions (e.g., relationships between

ideas).[27] Propositions represent underlying knowledge and include causal, conditional, temporal or Boolean (alternating and exclusive relations), algebraic (greater than) and categorical relationships. The number and quality of the propositions are representative of the underlying knowledge networks and can be compared to prototypical solutions or between groups of participants (e.g., novices vs. experts).[57, 63]

Semantic qualifiers analyze dichotomous signs and symptoms to identify four semantic classes: 1) reduced (few diagnostic features with limited linkages), 2) dispersed (extensive diagnostic features that are disorganized), 3) elaborated (extensive diagnostic features that are linked), and 4) compiled (rapid and correct summarization of the case). This method has been shown to be reliable and valid when compared to other measures of diagnostic competence (e.g., diagnostic accuracy, quality of thinking scores and global ratings of clinical reasoning and knowledge).[66] Experts use semantic qualifiers more frequently and in more diverse situations than novices, and diagnostic success is associated with elaborated or compiled classes.[67]

Semantic networks graphically represent relationships between ideas and propositions in knowledge networks and reasoning pathways.[57] Directed arcs connect nodes (e.g., clinical findings, pathophysiologic processes and diagnostic hypothesis). Forward arcs represent inferences that are made from the clinical data. Backward arcs start with a hypothesis then look back at the data to support the hypothesis.[63] Semantic networks identify the overall strategy the clinician uses to evaluate the data, the directionality of inferences and the coherence of the diagnostic explanation.[57] Coherence is established by examining the interconnections between the nodes and can be global (i.e., connections amongst all nodes without contradictions or loose ends) or local (i.e., consistency in a component of the clinical problem). Forward and backward

inferences can be weighted and enumerated to categorize the major method of forward or backward reasoning used to derive a diagnosis.[63]

Think-aloud protocols and opinion-on-reflection (e.g., recall, explanation protocols, diagnostic inventories) mostly assess analytical reasoning. Much of the non-analytical (System 1) processing is, by its very nature, unconscious and not available to verbalization.[39] The mere act of asking a clinician to verbalize a non-analytic process converts reasoning to an analytical (System 2) mode.[8] Visual reasoning is by its nature a completely non-verbal process. While some insight into visual reasoning can be gained by think-aloud protocols, they are unable to access much of the visual reasoning process as participants are not consciously aware of what they are doing. Further, the increase in time to verbalize the reasoning process interrupts the image analysis.[68] Thus, it stands to reason that instead of analyzing visual reasoning using verbal methods, technologies such as eye-tracking, will allow us to analyze the process in a higher fidelity manner.

2.4.3 Eye-Tracking

2.4.3.1 General

A subset of medical and veterinary specialties (pathology, radiology and dermatology) is highly visual and diagnostic reasoning relies heavily on perceptive skills and System 1 processing.[69] Eye-tracking technology allows a researcher to monitor the position of a viewer's eye to indicate visual attention and focus on key features of interest. Saccade and fixation points are not random but have a highly replicable "scan path" between viewing sessions.[70] Using eye-tracking technology, experts have been shown to initially globally and holistically view the image and then rapidly formulate a diagnostic hypothesis (non-analytic, System 1).[71] The

remainder of the viewing is systematic (analytic, System 2), aimed at finding visual features to confirm or dispute any number of diagnostic hypotheses.[72] The pattern of eye-movements produced by novices and experts also highlights differences in reasoning and knowledge organization in that novices demonstrate complex and disorganized patterns while experts demonstrate efficient and highly organized patterns.[69]

2.4.3.2 Eye-Tracking Studies in Human Radiology

Cumulatively, eye-tracking research in human radiology has highlighted several important findings pertinent to visual diagnostic reasoning. First, experts exhibit a higher density of fixation points on relevant than irrelevant areas of interest.[73] Second, using spatial frequency analysis there was shown to be a significant difference between lesion-containing areas that attracted visual attention and were correctly interpreted and those that were visually inspected but not reported.[74, 75] Third, radiologists had more fixations in cases where they agreed on how to diagnose the case than when they disagreed.[76] Fourth, true and false positive diagnoses were associated with longer dwell times than false negatives, while false negatives tended to have longer dwell time than true negatives.[77] This highlights important learning issues for radiologists, as they clearly visualized false positive and false negative locations, allowing for useful feedback in their learning. Fifth, false positive diagnoses yielded a different pattern of background scanning than false negative findings, biasing the further analysis of the image once a "diagnosis" was reached.[78] Sixth, during an expert's examination of a radiology image the initial 30% of the visualization time ("discovery search") is when an initial hypothesis(es) is formed. The remaining 70% is confirmatory scanning to corroborate or refute the initial hypothesis.[72] Seventh, perceptual learning in radiographic interpretation identified a higher

sensitivity to low contrast dots on radiographs in experts compared to novices and that novice sensitivity to low-contrast dots improved with practice.[79]

2.4.3.3 Eye-Tracking and Human Pathology

Two studies have utilized eye-tracking technology in human pathology.[6, 7] These studies identified similar features in pathology image interpretation as in other fields (e.g., radiology). They identified "sporadic" and "common" fields of view with pathologists, residents and medical students selecting 20%, 43% and 37% of sporadic locations respectively. In addition, pathologists (experts) were found to have significantly shorter total viewing times than residents and medical students.[6] From this study, two styles of slide scanning patterns have been identified for human pathologists, a "scanning style" and a "selective style". In the Tiersma et al.[7] (2003) study, however, a broad range of diagnoses were made and no association between scanning pattern or time to examine the slide was evident.[7] Eye-tracking and diagnostic reasoning is entirely new to the field of veterinary pathology. It is likely that lessons from human medicine are transferrable to veterinary medicine; however, without rigorous investigation, this is a working hypothesis.

2.5 Summary of Visual Diagnostic Reasoning Theory and Measurement

Experts in medical domains have been shown to differ from novices in three ways: 1) they spend proportionately more time establishing a representation of the problem before embarking on finding a solution,[80] 2) they rapidly assign the problem to a relevant and typically appropriate category prior to further processing,[81] and 3) the expert rapidly aligns their schemata to the particular aspects of the case.[82] Lessons from the human literature on visual diagnostic

reasoning suggest that differences in novices and experts are related to total time in which slides are viewed, selection of fixation points and time spent on fixation points. Using these three outputs, I aimed to investigate visual diagnostic reasoning in veterinary pathology using eye-tracking technology (see "Objective 1", Section 1.4).

2.6 Clinical/ Diagnostic Reasoning Education

2.6.1 Learning Theories

Veterinary students, as adults, have particular learning traits that are important in optimizing diagnostic reasoning learning. In adult learning, two key factors drive the learning experience: the learner and the context in which learning takes place.[83] Important to adult learning is that learners:

1) organize knowledge in structures described as "schemes", [83]

2) have a great deal of prior knowledge and new learning experiences should ideally build on this knowledge,[83, 84]

3) make sense of knowledge through active manipulation of information,[83]

4) have optimal learning when it is experiential and contextually and socially bound,[83]

5) are intrinsically rather than extrinsically motivated to learn,[83, 84] and

6) knowledge acquisition is hampered by excessive cognitive load.[85, 86]

2.6.1.1 Knowledge Structures

Key to adult learning is not only the knowledge that is acquired, but also the manner by which that knowledge is structured and subsequently accessed (recalled). Adult learners, as compared to children, have a wealth of prior knowledge gained through experiences and problem

solving which plays an important role in their learning. It is believed that the prior knowledge possessed by adults is organized in mental schemes (categorical memory structures).[83] During learning, schemes accumulate new knowledge that is tuned (gradually modified) or restructured (reorganized or new schemes are created to better represent a new understanding).[83] As well as adding, fine-tuning and re-structuring schemes, the ability of adults to meta-cognitively reflect on their mental schemes is important in the highest levels of evaluating and synthesizing knowledge. There is some suggestion that the structure in which students learn new material is also the structure used to recall material.[87, 88] This has important instructional design implications for teaching biomedical science in that students taught basic sciences in an organ-based manner will tend to recall their knowledge in the context of organs. This is disconnected from the way cases are presented to veterinarians and may hamper biomedical knowledge transference.[88]

Organization of knowledge is key to clinical veterinary expertise in that experts have more organized and detailed knowledge structures than novices.[11, 34, 37] It stands to reason that a goal of biomedical education is for novices to create or learn knowledge schemes that are similar to those used by experts.[89] During the process of learning, rather than ignore students' existing knowledge schemes, instructors can build on these schemes by correcting their structure or adding new material.[89] To do this, instructors need insight into what flaws are present within these existing knowledge schemata. An ideal way to achieve this is through classroom discussions with feedback on the understanding of the concepts from both the student and instructor. Another role of instructors or experts is to model their own knowledge structures. Many experts have abbreviated ways of understanding a concept or performing a task.[89] To transfer knowledge structures or understanding to a student, instructors need to be self-aware of which steps they skip and make explicit the links between the biomedical science and clinical

signs in a case.[58] The use of concept maps, algorithms and schemes are ways to provide students with cognitive structures and encourage deep learning.[89]

"Illness scripts" is an emerging theory of how knowledge is organized in experts and utilized during clinical reasoning. A comprehensive biomedical knowledge base is essential in illness script development and clinical problem solving.[87] Teaching biomedical science by emphasizing and making the causal links to clinical signs[90] and symptoms explicit improves retention of biomedical knowledge[91] and diagnostic success in difficult cases[92] (both achieved by developing the knowledge schemes in students). Teaching biomedical science in schemes that represent expert knowledge structures has been shown to enhance students' knowledge acquisition and problem solving capabilities.[87]

2.6.1.2 Prior Knowledge, Experiential and Active Learning

Fundamental to knowledge schemes is the prior knowledge that adults possess. In essence, the understanding of new knowledge and concepts cannot be isolated from prior experience.[83] Prior knowledge and experience is also identified as one of the key internal motivators for adult learning.[84] For experiential learning to be effective, new learning experiences need to be interactive and build on prior knowledge gained from earlier experiences (i.e., connect what is already known to the new experience).[83] These two features have important implications for veterinary education in that optimal adult learning experiences are contextual and active.

Key to experiential learning is that learning is an interactive process between the learner and the content. Knowledge can be viewed as a tool; it is only fully understood through its use.[93] Active learning in biomedical science is often discovery-based, uncovering new knowledge in

the context of how it will be used (in clinical medicine), and building on what the student already knows.[94] By engaging in a discussion about a biomedical problem, students relate the material to their prior experiences, re-assess their understanding based on others' opinions, and ultimately better understand and retain the knowledge.[94] It has been shown that biomedical science knowledge presented in a lecture format has a retention rate or about 30% where interactivity with cases increases the retention up to 75%.[95]

Models of experiential learning represent a cyclical process of abstract conceptualization, planning for implementation, active experimentation, concrete experience and reflective observation.[83] In the context of biomedical science education, the learner is introduced to a concept (e.g., abstract conceptualization in a lecture, text book or e-learning), provided with a concrete example of the concept (e.g., a clinical case illustrating the biomedical principle in small group problem or case-based learning session) and encouraged to engage in the problem (e.g., active experimentation).[96]

For most learning situations, particularly concepts that are difficult to understand, learning occurs not in a single cycle but in a spiral, where the concept is experienced multiple times with modifications and gains in comprehension with each passing stage of the learning process.[83] This has implications for instructional design and how it needs to be built on prior knowledge structures and understanding. Presenting similar concepts (disease processes as an example) and concrete experiences (cases) in increasing complexity allows the learner to build on their prior knowledge in an experiential manner.[96, 97] The role of the instructor in experiential learning is to relate new material to the prior knowledge of the learner, provide concrete experiences that students can actively engage in, and provide conceptual experiences in the form of narration for experiences that are difficult to deliver to a classroom setting.[83] An additional

role of the instructor is to create cognitive dissonance (in a safe environment) in students' understanding. This highlights the limitations of the students' current understanding, forcing students to re-examine and potentially restructure their knowledge schemes.[89] With respect to biomedical knowledge transference, the importance of the cycle of experiential learning is highlighted by data that a) if only a principle is taught, the likelihood of transfer is about 5%,[98] b) if the principle is illustrated with a single example it increases to 25%, but c) if multiple examples are provided transference can increase to 47%.[99]

2.6.1.3 Contextual Learning

Cognitive scientists argue that learning and cognition are fundamentally situated (i.e., contextually bound) and that ignoring this results in learning that is less robust and usable.[93] To create an analogy from learning languages, learning basic science as isolated facts out of context to their medical story is akin to learning words from a dictionary with no reference to sentence construction and use. Biomedical knowledge is situated within a clinical context and ignoring this context jeopardizes the effectiveness of the learning environment and successful transference of the knowledge. A counter-point to this argument, levelled against problem-based learning, is that knowledge gained within a context can become so bound to the context that transference to new situations is difficult.[83, 96] In terms of constructing contextual learning experiences, it seems that contextual learning is most effective once the general underlying biomedical principles are understood.[100]

2.6.1.4 Cognitive Load Theory

Cognitive load theory is an emerging theory that argues learning is largely restricted by the limitations of the working memory. The capacity of the working memory to recall and process new information is limited in that it is able to hold only 7 plus or minus 2 new pieces of information at any one time.[83] Working memory, however, has no known limit when processing information retrieved from the long-term memory. This has important implications for instructional design, in that the instructor is left with the task of identifying strategies to optimize the learner's memory available to structure and recall new information.

Three types of cognitive load have an impact on learning: 1) intrinsic load (the complexity of the learning task), 2) extraneous load (the manner which the learning tasks are presented) and 3) germane load (the load associated with actual learning).[85, 86] It is believed that these three forms are additive and, hence, educational strategies should be employed that maximize the germane load while minimizing the intrinsic and extraneous loads.[86] The intrinsic load (complexity) can be reduced by introducing multiple steps to problem solving, sequencing clinical cases from low to high levels of complexity, and by gradually increasing the fidelity of the case.[86] There are several approaches to reducing the extraneous load. These include providing a non-specific goal (e.g., to not to initially strive for a diagnosis, but to identify the biomedical principles at play in a case), introducing a worked through example of a problem, partially completing a problem, putting multiple sources of information in one location (e.g., temporally and spatially), replacing written explanations with visual aids (e.g., diagrams, images), and replacing multiple sources of the same information within one location.[85, 86] Germane load can be optimized by varying the cases (e.g., encouraging comparison between cases and identifying unifying principles), varying the types of cases encountered (e.g., mixed

practice instead of blocked practice), and self-explanation of concepts to increase prior-knowledge elements.[86, 96]

Based on learning theories, several teaching techniques have been developed, aimed at optimizing learners' knowledge, skills and attitudes in biomedical science. Unifying these teaching strategies is the aim of long-term retention and transfer of biomedical concepts into clinical and paraclinical practice.

In an educational context, students are believed to develop illness scripts during encounters with real patients and incorporate pre-learned biomedical knowledge through relating symptoms with relevant pathophysiologic knowledge networks.[33] Biomedical knowledge, as well as making sense of symptoms, further constrains scripts to acceptable attributes based on pathophysiological knowledge.[34] Novices transition through several knowledge structures as they develop towards experts. These include elaborated causal networks, abridged networks, illness scripts and instance scripts.[33] It is thought that the repertoire of illness scripts that clinicians acquire with experience allows problem solving by recognizing the presenting problem as being similar to ones previously solved (pattern recognition).[14] This builds on a cognitive psychology concept of heuristic judgment, the attribute-substitution model, where difficult questions are answered by substituting the answer to an easier question.[43] Attribute substitution, however, immediately introduces a systematic bias to judgment, particularly when the subject is faced with unfamiliar or unknown situations.[43] Attribute substitution provides initial input to many judgments; however, it can be supplemented, moderated or overridden by relevant logical rules and learned algorithms.

2.6.2 Teaching Clinical and Diagnostic Reasoning

Several themes emerge in the clinical reasoning education literature that builds on the cognitive underpinnings of illness scripts, and System 1 and System 2 reasoning processes. It is generally agreed that the most effective method for acquiring illness scripts and converting System 2 to System 1 reasoning is the repeated and directed exposure of students to real cases.[8-13] Rather than case exposure being a passive process, several educational principles are recognized: 1) learners need to be actively engaged, 2) new information is articulated on student's prior (biomedical) knowledge, 3) intermediate stages of clinical reasoning are explained, 4) learners use their clinical knowledge to judge clinical information to support or reject a hypothesis, and 5) immediate feedback from peers and teachers validates newly acquired knowledge.[13, 34, 88] Deliberate application and integration of biomedical knowledge is important in script building, placing biomedical knowledge in an accessible clinical-presentation based location.[35, 37] Teaching the basic sciences in a clinical context facilitates the encapsulation of biomedical knowledge in illness scripts.[37] Both increased case reasoning exposure and practice, as well as the teaching of biomedical knowledge in the context of clinical cases, are supported by newer medical curricula including case-based and clinical-presentation-based curricula.[48, 101, 102] The teaching of knowledge "schemes" has been suggested to enhance memory organization and diagnostic success.[14] Over-reliance of novice diagnosticians on non-analytical (System 1) approaches can lead to diagnostic error.[12] It is, thus, important that novice diagnosticians need to acquire affective de-biasing strategies.[8, 42] These include instructions to explicitly list evidence in the case (i.e., promoting analytical reasoning),[12] awareness of known cognitive biases and adverse effects on the diagnosis, forced entertainment of alternate possibilities,[88] metacognitive reflection of own thought processes, reduced reliance on memory (e.g., through use of

mneunomics, algorithms, hand held computers), and use of strategies to avoid predictable biases.[103, 104] Finally, use of multiple reasoning processes (providing redundancy) lead to the most success (i.e., accuracy) in diagnostic reasoning.[10, 11, 105] Analytic and non-analytic reasoning processes likely occur interchangeably and additively in a reasoning process.

2.6.3 Teaching Visual Reasoning

Like many medical and veterinary specialties, veterinary clinical pathology is most commonly taught in an unstructured, apprenticeship manner. Expertise is attained through long, repetitive and supervised practice. With this, novices gradually acquire a repertoire of visual patterns, and those patterns are linked to diagnoses, disease outcomes, expected disease behaviours, pathogenesis, and confirmatory testing modalities. This builds on educational principles associated with acquiring visual expertise in that the acquisition of visual images (e.g., scripts) are acquired through the controlled exposure of novices to multiple cases.[13] In the undergraduate DVM programs, pathology (cytology) images are often taught in a didactic manner with static images. In practice, images on a microscope are fluid; thus, there is a disconnect between how visual reasoning is taught and practiced. Building on the research on the impact of context on clinical reasoning,[52] it is possible that this disconnect in teaching and practice may affect the ability of a student to transfer their visual knowledge into a clinical context.

A long-term difficulty associated with teaching highly visual skills, is an inability of the instructor and student to be sure they are looking at the same thing. It is assumed that what the instructor is referring to on a histopathology or cytology slide (with the exception of arrows) is what the student is looking at. Arrows, built into the microscopic visual field, are only useful to

demonstrate specific features, and the subtler, but more diagnostically useful patterns are less demonstrable. Similar to the identification of features or cues in a clinical situation, pathologists reading microscopic pathology slides identify key diagnostic areas.[6] Due to the complex nature of the visual task, and previous poor accessibility of cytology to other forms of visual reasoning measurement (e.g., eye tracking data), very little research has focused on how to best teach visual reasoning.

Research in multi-media learning (i.e., how people learn and think using graphic images) has highlighted three general educational principles that can be applied to veterinary pathology visual diagnostic education.[106] Firstly, learners are more successful at learning from graphics when the relevant features are highlighted on the image, reducing the cognitive load.[107, 108] Secondly, if a learner has a high prior knowledge, they learn better from graphics compared to those with low prior knowledge.[73] Thirdly, learning is enhanced when a spoken text, rather than printed text accompanies the graphics.[109] This suggests that use of these techniques may also be effective in teaching visual reasoning in veterinary clinical pathology.

2.7 Research Question

Is there a quantifiable difference in visual diagnostic reasoning performance between expert and novice veterinary pathologists using eye-tracking technology? Does the introduction of two educational interventions facilitate visual diagnostic reasoning development in novice veterinary pathologists?

2.8 Objectives of the Study

1. To quantitate visual diagnostic reasoning performance in expert and novice veterinary pathologists
    a. To investigate the use of eye-tracking methodology in measuring visual diagnostic reasoning skills by comparing it to established think-aloud protocols and measures of diagnostic success.
    b. Novice and Expert studies:
        i. To use eye-tracking technology as a measurement tool for visual diagnostic image evaluation and to establish baseline differences in eye-tracking ability in novice vs. expert pathologist.
        ii. To investigate dual process theory as a potential continuum between novice and experts based on think-aloud recordings.
2. Visual reasoning educational interventions.
    a. To determine if the use of explicit features in image interpretation improves novice pathologist visual diagnostic reasoning skills.
    b. To determine if the introduction of pathology case repetition improves novice pathologist visual diagnostic reasoning skills.

Chapter Three: **Methods**

3.1 Eye-Tracking

3.1.1 Equipment

Applied Science Laboratory (ASL) Mobile Eye-XG® eye-tracking glasses, portable Data Transmit Unit (DTU) and EyeVision® software were used for the collection of eye movement data for both the novice-expert (Objective 1) and visual reasoning educational interventions study (Objective 2) (ASL, Bedford, MA). The eye-tracking data, environment mapping, area of interest selection and data analysis were collected using EyeVision® software and analyzed using GazeMap® software (Applied Science Laboratories, Bedford, MA). The images were projected on to a 40-inch 1080p LCD high definition (HD) screen (Toshiba 40E21OU ®) with the participant, wearing eye-tracking glasses sitting one metre from the screen (Appendix 1).

3.1.2 Image Selection

3.1.2.1 Image Capture

Five different subcutaneous tumors from dogs were used for both the novice and expert (Objective 1) and visual reasoning education interventions (Objective 2) studies. All five tumors belong to the "round or discrete" cell tumor group, and are the top five major differential diagnoses for subcutaneous round cell tumors in dogs. The tumors were: 1) cutaneous histiocytoma (CH), 2) cutaneous lymphoma (LSA), 3) transmissible venereal tumor (TVT), 4) mast cell tumor (MCT), and 5) extramedullary plasmacytoma (PCT). This group of tumors was chosen as subcutaneous mass aspirates are commonly submitted to clinical pathologists in diagnostic laboratories, and the diagnostic features of these tumor groups are taught to veterinary students during their DVM studies.

Fine needle aspirate preparations from five subcutaneous masses were used. Aspirates were all collected by veterinarians at referring veterinary practices and submitted to Antech Diagnostics, Calgary. The cytologic preparations were air-dried and stained with a Wrights-Giemsa modified stain using a Hematek® automatic stainer. Images were captured by a board-certified clinical pathologist (Amy Warren) using a Nikon Eclipse Ni® microscope, Nikon DS-Fi2 camera® and Nikon NIS Elements BR® program (Images 1, 5, 6, 7, and 9). The regions of the cytologic preparations where photographic images were taken were selected for 1) diagnostic representativeness, 2) distribution of diagnostic features, and 3) image quality. Additional images were sourced from Noah's arkive® free-source educational images (Images 2, 3, 4, 8 and 10).

The same ten images of the five different subcutaneous round cell tumors (duplicates or parallel samples of each type of tumor) were used for both the novice and expert (Objective 1) and visual reasoning education interventions (Objective 2) studies (Figure 3.1a and 3.1b). The duplicate images were randomly assigned into the two sets of slides (slides 1-5 and 6-10), with one image of each tumor type represented in each set. The images were: 1) cutaneous histiocytoma (images 1 and 6), 2) cutaneous lymphoma (images 2 and 10), 3) transmissible venereal tumor (images 3 and 8), 4) mast cell tumor (images 4 and 9), and 5) extramedullary plasmacytoma (images 5 and 7).

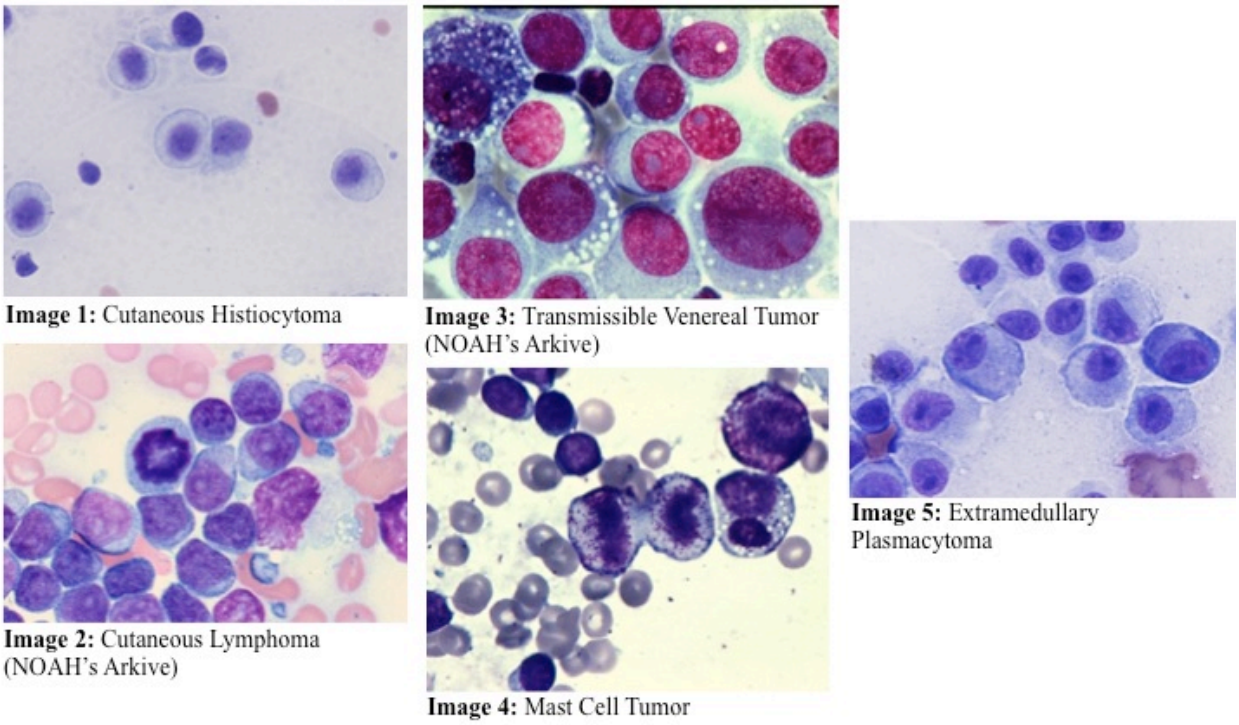


Image 1: Cutaneous Histiocytoma

Image 3: Transmissible Venereal Tumor (NOAH's Arkive)

Image 5: Extramedullary Plasmacytoma

Image 2: Cutaneous Lymphoma (NOAH's Arkive)

Image 4: Mast Cell Tumor

Figure 3.1a Images 1 to 5 used in novice and expert (Objective 1) and visual reasoning educational interventions studies (Objective 2).

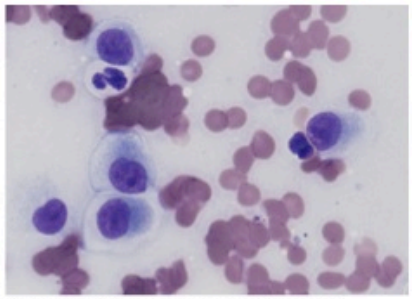


Image 6: Cutaneous Histiocytoma

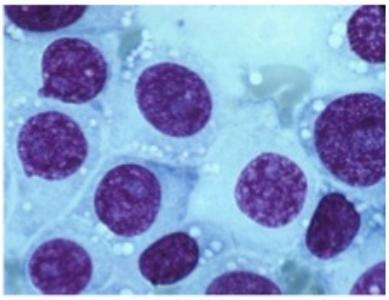


Image 8: Transmissible Venereal Tumor (Noah's Arkive)

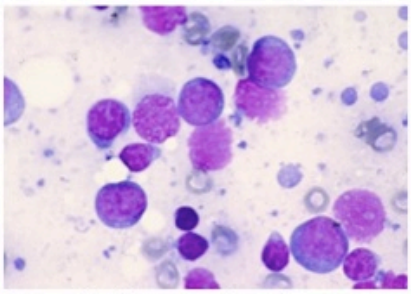


Image 10: Cutaneous Lymphoma

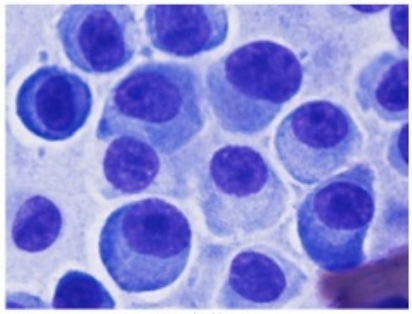


Image 7: Extramedullary Plasmacytoma

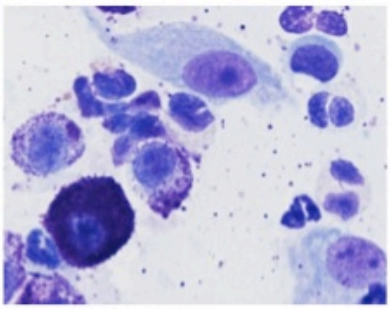


Image 9: Mast Cell Tumor

Figure 3.1b Images 6 to 10 used in the novice and expert (Objective 1) and visual reasoning educational interventions (Objective 2) studies.

3.1.2.2 Image Validation

The ten images of the five subcutaneous round cell tumors were viewed independently by three board-certified veterinary clinical pathologists (Dr. Amy Warren BVSc, Dipl ACVP; Dr. Catherine Wagg DVM, Dipl ACVP; and Dr. Heather Preist DVM, Dipl ACVP). There was complete 100% agreement on the diagnosis for all 10 images.

3.1.3 Image Backgrounds and Areas of Interest

3.1.3.1 Image Environment Map Creation

To allow eye-tracking to be performed and to establish trackable areas of interest (AOIs) on each of the ten round cell tumor images, an environment map for each image was created using the MobileEye-XG® eye-tracking glasses, EyeVison®, and GazeMap® software. Briefly, a slow motion recording of each of the images projected on the 40 inch, 1080p LCD HD screen was created with an environment reference target established at the top left hand corner of the screen (i.e., this target allows the software to create a map of the scene based on visually prominent points) (Appendix 2). Using this recording, an environment map for each of the 10 images was created using the GazeMap® software.

3.1.3.2 Area of Interest (AOI) Selection and Validation

For each image, three to five AOIs were identified that represented key features in the image that were important to visualize in order to reach a correct diagnosis ("diagnostically useful" areas) (Appendix 3a and 3b). The AOIs were identified *a priori* and were based on the published key features indicative of the five tumors selected for this study.[110] Further validation of the representativeness of the AOIs selected for each image came from the think-aloud data

from the novice and expert experiment (Objective 1, Chapter 4) where the verbalized key diagnostic features elicited during the think-aloud protocol from the experts, corresponded with the cell features present in the selected AOIs.

Using GazeMap® software, each AOI was drawn from multiple angles of vision on the created environment map. The tracking of the AOI over a range of visual angles on the environment map was also tested. Drawn AOIs were checked visually by two observers (Amy Warren and Jason Abboud, research assistant). To ensure that the visual recording performance of the participants during live eye-tracking of the AOIs (i.e., to be sure that when a participant was visualizing an AOI, this was being detected by the program), each AOI for each image was visualized without a break for 5 seconds consecutively by a non-participant viewer (Jason Abboud). The data collected were then analyzed using GazeMap® and AOIs were accepted when greater than 95% of the total time was spent in the AOIs when AOIs were visualized consecutively (i.e., the remaining 5% or less of the recorded time were gaps in eye-tracking due to participant's blinking or being distracted).

3.1.4 Participant Calibration and Validation

For each of the participants, it was necessary to calibrate the eye-tracking equipment to their eyes, prior to data recording. This personalized calibration was performed prior to each eye-tracking recording session. Briefly, the eye-tracking glasses were fitted firmly to each participant, such that with head movement there was no movement of the glasses with respect to head rotations from left to right or up and down. The "pupil" camera mounted on the glasses was adjusted so that the pupil was clearly visible, and so that the three tracking lights were focused and visible over a range of eye-movements. The "scene" camera was adjusted such that

the LCD screen and projected cytology slide image were in full view. The two recordings from the "pupil" and "scene" cameras were merged by the Mobile-eye-XG software to track the eye-movements of the participant, with a red coloured cross hairs moving across the scene as the pupil position changed. The Mobile-eye-XG software was then "aligned" and calibrated to the participant's eye. To allow the computer software to match the pupil recordings with the scene recordings, the participants were asked to visualize 12 points on the LCD screen and projected image, while an operator (Amy Warren or Jason Abboud) correspondingly recorded the 12 point locations using the Mobile-eye-XG® software. The 12 points selected *a priori* were 8 numbers situated around the periphery of the screen and four figures on the projected "calibration" slide (i.e., a star, heart, arrow, and smiley face) (Appendix 2). To ensure that the calibration was successful, each participant was asked to maintain visual contact with each of the four figures (i.e., a star, heart, arrow and smiley face) for 5 consecutive seconds. Using GazeMap®, the amount of time the participant was viewing the four AOI calibration figures was calculated (e.g., with the potential to achieve up to 100% of the total 20 second recording time). When greater than 90% of the participant's view time was recorded within the AOIs, the calibration was accepted. Each recording was also reviewed by the calibrator to ensure that the remaining 10% or less of the time was due to participant blinking, eye-movement between images or distracted visualization away from the figures or screen.

3.1.5 Data Collection and Video Recording

For both studies, data collection for the eye-tracking phases were recorded using the Mobile-eye-XG and analyzed using the GazeMap® software programs. Participants were instructed that all of the images were fine needle aspirates from subcutaneous masses from dogs,

and that they had a maximum of 2 minutes to view the image and come up with a diagnosis. If the participants were unable to reach a diagnosis, the participants were to state "no diagnosis". Once a diagnosis was reached the image was removed immediately from the LCD screen. The eye-tracking movement and time was recorded from the moment the image was shown to the participant on the screen and ended as soon as the participant had a diagnosis. The image was then re-shown to the participant and they were asked to re-iterate the diagnosis and identify the key visual features that lead them to the diagnosis. The diagnosis and key visual features identified were recorded manually. To ensure accuracy of written data collected, the entire session was recorded using a webcam facing the participant and a Sony video recorder facing the image screen.

3.2 Objective 1: Novice and Expert Study

3.2.1 Research Design

A pre-experimental, static group comparison was used with four groups: Novice 1, Novice 2, Expert 1, and Expert 2. Novices and experts were randomly divided into groups 1 (Novice 1 and Expert 1) and 2 (Novice 2 and Expert 2). Group 1 participants (Novice 1 and Expert 1) were shown slide images 1 to 5 first, while concurrently using the "think aloud" protocol and viewing the images with the eye-tracking glasses on. This was followed by slide images 6 to 10 with the use of eye tracking but without using think-aloud (Figure 3.2). For group 2 participants (Novice 2 and Expert 2), this order was reversed (Table 3.1).

Table 3.1 Novice and Expert study design and data collection. Eye-tracking (ET) and Think-aloud (TA) protocol Images 1 to 5. ET only Images 6 to 10.

Group	Part 1	Part 2
Novice 1 (n = 5)	ET + TA Images 1-5	ET only Images 6-10
Novice 2 (n = 6)	ET only Images 6-10	ET + TA Images 1-5
Expert 1 (n = 4)	ET + TA Images 1-5	ET only Images 6-10
Expert 2 (n = 2)	ET only Images 6-10	ET + TA Images 1-5

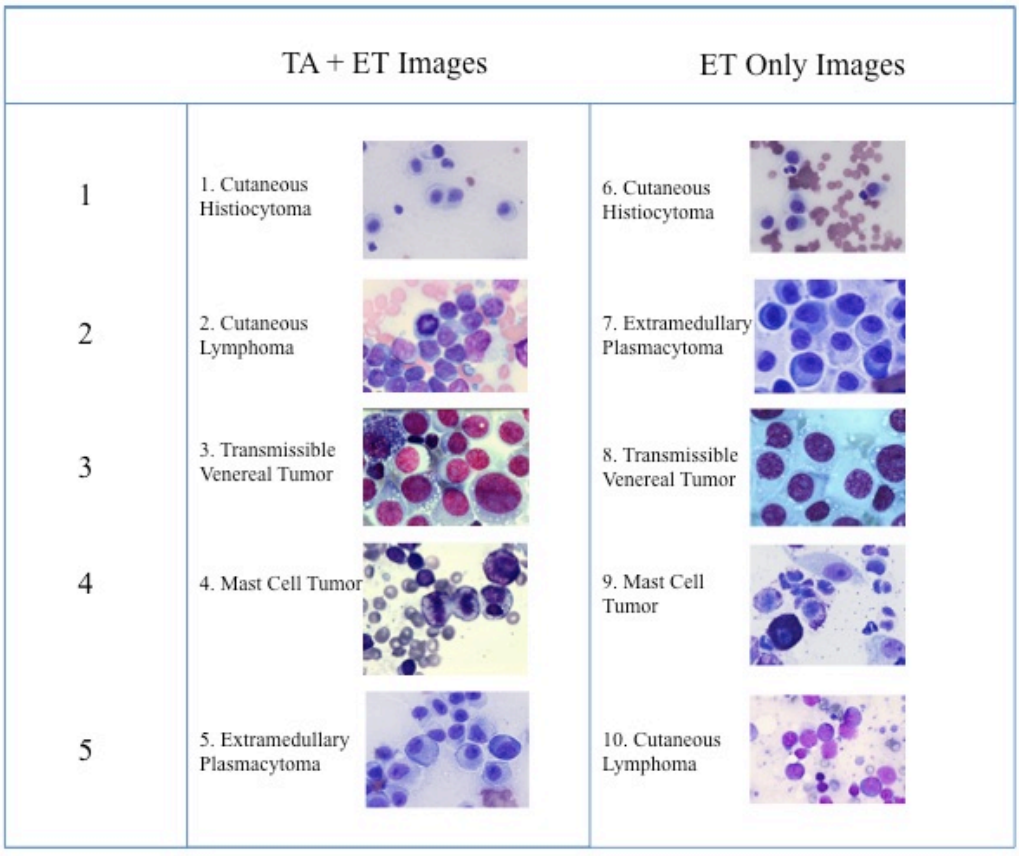


Figure 3.2 Image order for Novice and Expert study based on when the think aloud (TA) and eye-tracking (ET) protocol (Images 1 to 5) and ET only (Images 6 to 10) were used.

3.2.2 Participants

Experts were recruited by invitation to participate voluntarily in the novice and expert study. Due to the small number of experts in this field, experts were recruited from both the University of Calgary (n = 1) and Cornell University in New York, USA (n = 5). Experts were all practicing diagnostic veterinary clinical pathologists. To meet the requirements for an expert, participants had to be qualified DVMs, registered for practice in Canada or the United States, had complete post-DVM residency programs in veterinary clinical pathology (3 years), and were Diplomates of the American College of Veterinary Pathologists (Dipl ACVP).

Concurrently, novices from the University of Calgary were also recruited by invitation to participate voluntarily in the novice and expert study. The novices were a combination of DVM students (5/11, 45%) that had completed one (n = 2) or two years (n = 3) of DVM training, and DVM graduates enrolled in graduate studies (6/11, 55%). All of the DVM graduates were categorized as novices as they had never practiced veterinary medicine, were completing masters and PhD studies in fields other than veterinary pathology, and had no further experience in veterinary pathology than what they attained during their DVM studies. Five of the six DVM graduate novices had DVM degrees that were not recognized for registration in Canada.

Prior to data collection, participants were asked to fill out a short questionnaire collecting demographic information and previous experience in cytology or histopathology (Appendix 4). The participants were also asked to sign a consent form approved by the Conjoint Health Research Ethics Board (Ethics ID 24239) from the University of Calgary (Appendix 5).

3.2.3 Projected Images

The combined slide images (Figure 3.2) were put into two different Powerpoint presentations. One set for use with the Novice 1 and Expert 1 groups where the think aloud protocol and eye-tracking for images 1 to 5 initially, were followed by the use of eye-tracking only for images 6 to 10. The second set for use with Novice 2 and Expert 2 groups where eye-tracking only images 6 to 10 were initially introduced followed by the think aloud protocol and eye-tracking for images 1-5. All slide images were projected onto a 40-inch, 1080p LCD HD screen.

3.2.4 Think Aloud Protocol

3.2.4.1 Think aloud protocol script

During the think-aloud section (images 1 to 5), participants were given the following instructions: *"You will be asked to view a number of stained cell slides that are fine needle aspirates from subcutaneous masses in dogs. Each slide will appear once we have synchronized the eye-tracking headset and software program so that we can establish a baseline and begin the process of collecting your data."* Prior to showing the participants images 1 to 5, a think aloud "practice slide" was shown (Figure 3.3) and the following instructions given: *"I will begin by showing you a practice slide of "Where's Waldo?" followed by a set of 5 stained cell slides, one at a time. In particular, I need you to verbalize your thinking about what you see on the slide as soon as it becomes visible on the TV monitor. I need you to talk about what it is you are thinking as you view specific or general features of the slide that lead you to a clinical diagnosis (or to locate Waldo)."* The participants were asked if they understood the instruction and any of their questions or

concerns were clarified. Initially, the "Where's Waldo?"[2] practice slide was shown to establish the think aloud protocol and procedures. If participants were silent for more than 5 seconds, they were prompted by the researcher: *"If you could keep verbalizing what you are thinking about this slide."* After the "Where's Waldo?" practice slide established that the participants understood the protocol expectations, images 1 to 5 were then shown one slide at a time. Once a participant had sufficient time to view the image (note: a maximum of 2 minutes was established *a priori*, however, not one participant required this amount of time) to talk through the slide and derive at a final diagnosis or was unable to decide (i.e., recorded as no diagnosis) the eye tracking was stopped, and the image removed. The image was then reshown to the participant and they were asked: *"Based on your clinical diagnosis for this slide (provide participant with their response), how did you derive this decision."* The diagnosis for the slide as well as the responses elicited from the participant to this question was recorded manually and confirmed later on video review. After the participant had finished talking, the researcher would state: *"Good, we will now go on to the next slide."*

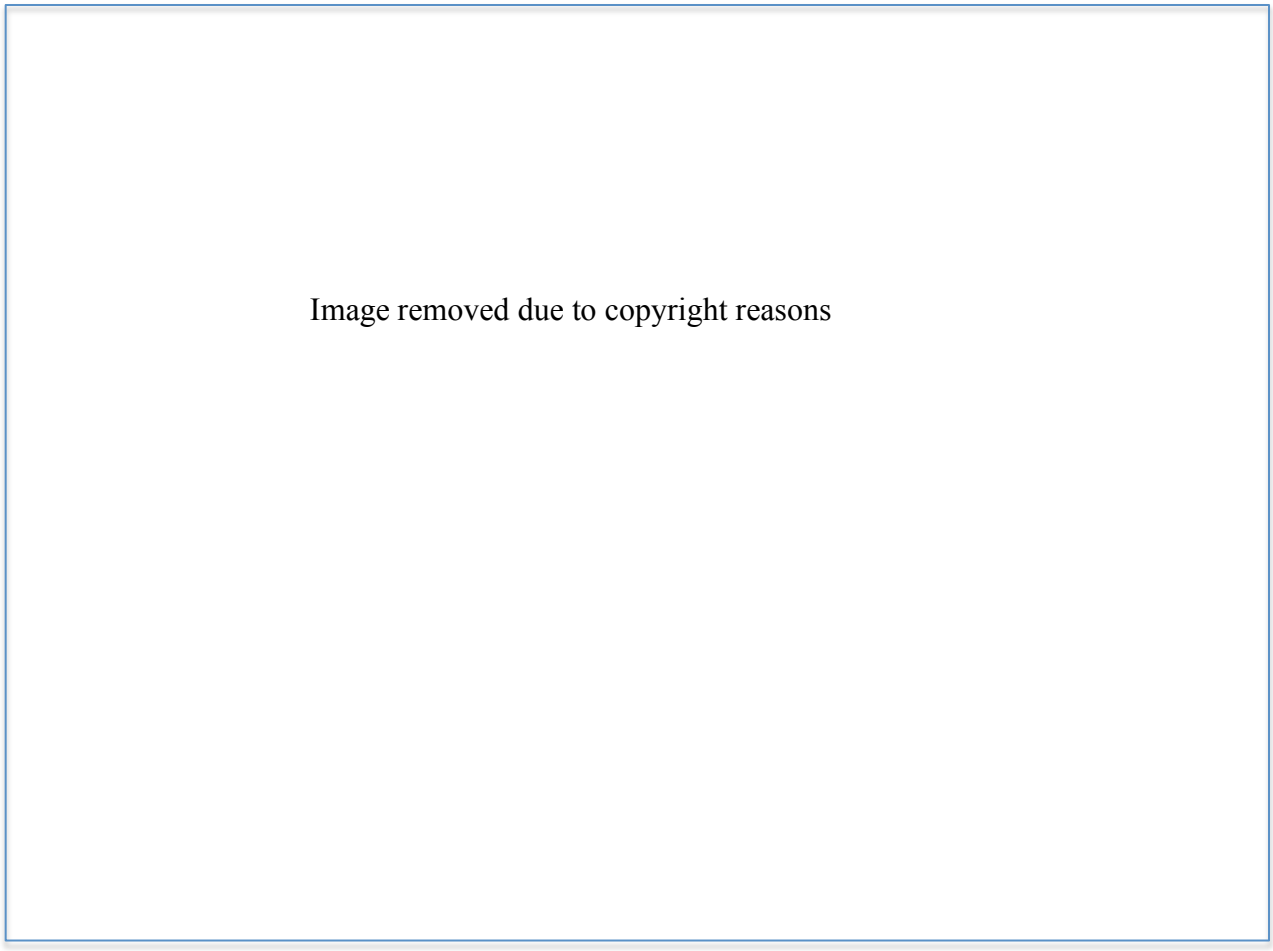


Figure 3.3 "Where's Waldo?" practice think-aloud slide.[2]

For the eye-tracking only part of the data collection process (Images 6 to 10) the participants were instructed as follows: *"You will now be asked to view another set of stained cell slides. Unlike the first set of 5 slides, I am going to ask that you simply focus on viewing any specific or general features of the slides without verbalizing or talking out loud. Once you think you have arrived at an appropriate clinical diagnosis, tell me what that is and we will stop recording."* After a diagnosis was reached or the participant was unable to decide, the eye-tracking was stopped, the image was removed and the participant was asked to review the slide prompted by: *"Based on your clinical diagnosis for this slide (provide participant with their response), how did you derive this decision."* The diagnosis

and responses elicited were manually recorded. Diagnoses for both sections of the experiment were graded as correct (1) or incorrect (0).

3.3 Objective 2: Visual Reasoning Educational Interventions

A quasi-experimental, pretest-posttest comparison group research design was used where participants were divided into 3 groups: Group 1) traditional teaching intervention group (comparison group taught in a traditional, didactic manner; $n = 8$); Group 2) basic visual reasoning teaching intervention group (explicit features and single image; $n = 10$); and Group 3) extended visual reasoning teaching intervention group (explicit features and image repetition; $n = 10$). For all three groups, the same pre-test using five slide images (1 to 5) (Figure 3.4) of the five subcutaneous round cell tumors (i.e., cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma,) was completed. This was followed by an hour-long teaching session using the three different teaching intervention strategies (Appendix 6). Immediately after the teaching session, a post-test using the five different images of the five subcutaneous round cell tumors (images 6 to 10) was performed (Figure 3.4). All pre- and post-testing was performed at one time (i.e., before and after the teaching intervention, respectively) to reduce the maturation effect between participants (Table 3.2).

Table 3.2. Visual reasoning educational interventions study design.

Teaching intervention group	Pre-test	Teaching session (1 hour)*	Post-test
1: Traditional teaching intervention: didactic	Images 1-5	Didactic powerpoint presentation with no active learning	Images 6-10
2: Basic visual reasoning teaching intervention: active learning, single image	Images 1-5	Didactic powerpoint presentation followed by active learning session with a single imag of each of the 5 round cell tumors	Images 6-10
3: Extended visual reasoning teaching intervention: Active learning, image repetition	Images 1-5	Short didactic powerpoint presentation with active learning using 5 different images of each of the 5 round cell tumors (2 images in total)	Images 6-10

*Detailed lesson plans for teaching interventions in Appendix 6.

3.3.1 Participants

Twenty-eight, final year (i.e., fourth year) Doctor of Veterinary Medicine (DVM) students from the University of Calgary participated in the study. All of the participants had the same educational background in veterinary clinical pathology and were enrolled in the fourth year clinical rotation in diagnostic veterinary pathology at the time of data collection. The students were divided into three groups on the basis of the timing of their rotation. Rotation groups were assigned based on external factors (i.e., predominantly based on the fourth-year rotations selected earlier in the academic year by the student). The participants for this study were different to the participants in the novice and expert study (Objective 1). Prior to the pre-

test, all participants were asked to sign a consent form approved by the Conjoint Health Research Ethics Board (Ethics ID 24239) from the University of Calgary (Appendix 5).

3.3.2 Pre- and Post-Tests

3.3.2.1 Images for pre- and post-test eye-tracking

The same 10 slide images used for the novice and expert study were also used for the visual reasoning educational interventions study. Images 1 to 5 were used for the pre-test and images 6 to 10 were used for the post-test (Figure 3.4). The images were put into two different Powerpoint presentations, one with the pre-test images 1 to 5 and the second with the post-test images 6 to 10. All images were projected onto the 40-inch, LCD screen (Appendix 1) in an identical set up to the novice and expert study.

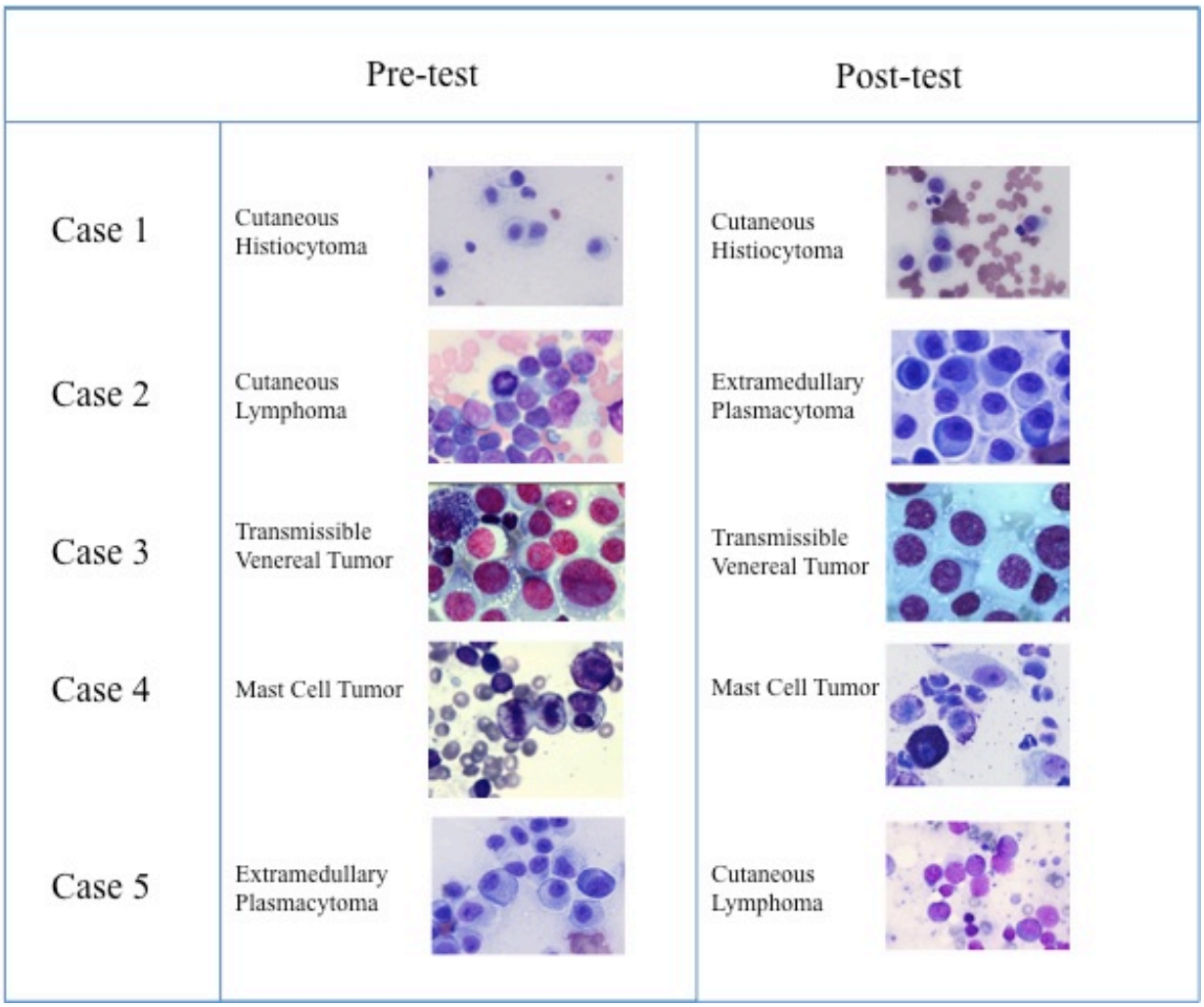


Figure 3.4 Image order for visual reasoning educational interventions study for the pre-test and post-test.

3.3.2.2 Eye-tracking

For both the pre- and post-tests, participants were instructed that they would be shown 5 images of fine needle aspirates from subcutaneous masses in dogs. For each slide, they were to visualize the image and come up with a diagnosis within a maximum of 2 minutes viewing time. Eye-tracking started as soon as the image was visible on the LCD screen. Once a diagnosis was

reached, the eye-tracking was stopped and the image removed. The image was then re-shown to the candidates and they were asked what key features they were viewing that led them to the diagnosis. The diagnosis and key visual features were recorded manually. Diagnoses were marked as correct (1) or incorrect (0). All pre- and post-test sessions were recorded on the web-cam and Sony video recorder.

3.3.3 Teaching sessions

3.3.3.1 Traditional teaching intervention (Group 1: Didactic teaching)

Students in this group were taught in a didactic manner using Powerpoint (Appendix 6a). The Powerpoint presentation took an hour and covered topics in the following order: an introduction to the types of subcutaneous masses in dogs, an algorithm for the cytologic diagnosis of canine subcutaneous masses, an algorithm for the 5 different round cell tumor types in dogs, visual features used to differentiate the 5 round cell tumors, the prognosis of the 5 round cell tumors, and the medical and surgical management of the 5 round cell tumors. Students were able to ask questions during the session but no active participation was done. The teaching session was recorded by video. This session was taught to two student rotation groups on two different dates.

3.3.3.2 Basic visual reasoning teaching intervention (Group 2: Active learning, single image)

Students in Group 2 actively participated in the learning session. They were asked to diagnose, using key features for each of the 5 round cell tumors described in the teaching session, a single representative image for each of the five subcutaneous canine round cell tumors (Appendix 6b). Prior to the active learning, the students were given a Powerpoint presentation

(20 minutes) that covered the following topics: introduction to the types of subcutaneous masses in dogs, an algorithm for the cytologic diagnosis of canine subcutaneous masses, the 5 subcutaneous round cell tumors in dogs, and visual differentiating features of the 5 subcutaneous round cell tumors in dogs (i.e., the features taught were based on clinical experience, current text book descriptions of cytological features of the tumors,[110] and the diagnostic features elicited from the novice and expert study). After the initial teaching session (30 minutes), the students from each group were given 5 images of the 5 different subcutaneous round cell tumors and instructed to diagnose the images based on the visual differentiating features provided in the Powerpoint. In addition, they were also explicitly told to compare and contrast the cases (20 minutes). The final 10 minutes of the teaching session were a review of the 5 cases with the instructor (Amy Warren) to ensure that the: a) correct diagnosis was reached, and b) students were able to identify the visual diagnostic features that enabled them to reason through to a diagnosis.

3.3.3.3 Extended visual reasoning teaching intervention (Group 3: Active learning, image repetition)

Students in the extended visual reasoning teaching intervention group also actively participated in the learning session, however, this group had 5 different images of the 5 different subcutaneous round cell tumors to diagnose (i.e., 25 images in total) (Appendix 6c). Prior to the active learning, the students were given a short Powerpoint presentation (10 minutes) that covered the following topics: the 5 subcutaneous round cell tumors in dogs and visual differentiating features of the 5 subcutaneous canine round cell tumors. The students were then given 25 images of the 5 different subcutaneous round cell tumors (5 of each tumor) that were

presented in a mixed practice model (35 minutes). They were instructed to use the key features to diagnose each of the images and specifically to compare and contrast the cases. After this, the instructor spent the final 10 minutes reviewing the 25 cases to ensure the students had correctly diagnosed the cases and were able to identify the visual diagnostic features.

3.3.3.4 Images used for teaching sessions

For teaching intervention group 1, the cytology images used in the teaching session were all from the pre-test (images 1, 2, 3 and 5), with the exception of the mast cell tumor where the pre-test image (image 4) was used as well as an additional image. For teaching intervention groups 2 and 3, the pre-test images were used in the Powerpoint presentation section and new images were used for the active learning session. For group 3, the tumor images used in the active learning section were randomized such that students were able to do mixed, rather than blocked practice. All images were captured from cases submitted to Antech Diagnostic services by the researcher. The post-test images were never used in the teaching sessions.

3.4 Data Analysis

3.4.1 Eye-Tracking

Gaze-map software was used to analyze the eye-tracking recordings for each session. For all eye-tracking recordings, except for three images in the novice and expert study, a percentage confidence in subject camera recordings of greater than 98% was attained (i.e., this indicates the percentage time that the gaze-map software was able to match the pupil movement with respect to the scene). The images where greater than 98% confidence in subject camera recordings was not possible were in the novice and expert study. In particular, for expert 4 on image 4 a 96.2%

percent confidence (1.46 second interval) and image 9 a 96.2% percent confidence (0.13 second) was recorded, and for expert 5 on image 9 a 97.9%percentage confidence (1.1 seconds) was recorded. A percentage subject confidence of greater than 98% was attained for all recordings for the visual educational interventions study (Objective 2). After a percentage subject confidence of greater than 98% was established, statistical analyzes for each of the tracking events (for each slide, for each participant) were calculated using the Gaze-map software. The data reported by the Gaze-map software for the total viewing included: total length of recording (s) and (frames), first AOI to fixate on, percentage time spent in any AOI, percentage time spent outside any AOI, and percentage time spent outside the image. For each AOI, the Gaze-map software also calculated the number of gaze points in the region, time in the region(s), number of dwells, average dwell duration(s), time to first gaze(s), percentage gaze in region (specific to each AOI) to total gaze time, percentage gaze in region of any AOI, number of fixations, average duration of fixations(s), total fixation duration(s), time to first fixation(s), percentage fixation time to dwell time, average pupil diameter in region (pixel), and *SD* of pupil diameter in region (pixel) (Appendix 7).

3.4.2 Think-aloud

For Objective 1, both a concurrent (slide images 1 to 5) and retrospective (slide images 1 to 5 or 6 to 10) think aloud protocols were used. For the retrospective "think-aloud" participants were asked, after viewing the image and coming up with a diagnosis, *"Based on your clinical diagnosis for this slide (provide participant with their response), how did you arrive at this decision."* These responses were recorded manually and, typically, the participants listed key visual features of the slide. The concurrent think-aloud recordings were reviewed and key

phrases related to the diagnostic reasoning (*"I am trying to see if there are any features that make me remember something"*), diagnostic features (*"large lymphocytes", "nuclei with two darker patches"*) or the diagnosis were recorded.

The visual features listed by the experts in concurrent and retrospective analyses were tabulated as a "gold standard" of diagnostic features for each of the slide images. For each of the experts, diagnostic features they identified were tabulated. The novices' concurrent and retrospective think-aloud protocols were analyzed in comparison to the experts' identified key feature responses. When the novices' responses were the same as the experts', these were tabulated for frequency comparisons. Additional features identified by the novices that were incorrect or irrelevant to the diagnosis were also recorded and tabulated. Non-diagnostic observations by the novices were recorded manually (Appendix 8).

3.4.3 Statistical Analysis

The statistical analysis was performed using SPSS version 20®. Because multiple tests were run, a Bonferroni adjusted *p*-value of 0.017 was used. Descriptive statistics for each group in the novice and expert study (Objective 1) (i.e., groupings of novice 1, novice 2, novice overall, expert 1, expert 2, and expert overall) and the educational interventions study (Objective 2) (each slide on the pre- and post-test for each educational intervention group as well as pre- and post test results aggregated) included the mean (*M*), standard deviation (*SD*), kurtosis and skewness values. Cases excluded from data analysis as a result of eye-tracking recording errors in Objective 2 (educational interventions) were pre-test slide 1 for student 3, post-test slides 1 to 5 for student 13, and post-test slides 3 to 5 for student 2. The corresponding, parallel pre-test and post-test slides for each of the paired diagnoses for the excluded slides were also excluded from

analysis. Diagnostic accuracy scores from each of the participants whose eye-tracking data were excluded were included in the analysis as these data were complete. No cases were excluded from the novice and expert study.

An independent samples *t*-test was used to analyze between group differences for novice 1 vs. novice 2 and expert 1 vs. expert 2 groups, for comparing eye-tracking only vs. think aloud and eye-tracking slides and to compare novices aggregated vs. experts aggregated for time to diagnosis and the percentage time spent in the AOIs. A chi-square test was used to analyze differences in diagnostic accuracy for between group differences for both the novice-expert (Objective 1) and the educational interventions (Objective 2) studies. For Objective 2 (educational interventions) paired sample *t*-tests were used to compare pre- and post-tests for changes over time for the time to diagnosis and the percentage time spent in the AOIs. A one-way analysis of co-variance (ANCOVA) was used to compare the different teaching groups' performances pre- and post-test for the time to diagnosis and the percentage time spent in AOIs for individual diagnoses as well as the results from the diagnoses aggregated.

For both objectives, a Cohen's standard *d* was calculated for effect size differences for parametric analyses.[111] A *d* of 0.20 to 0.49 was considered a small effect size, a *d* of 0.50-0.79 was considered a medium effect size, and a *d* of $\geq$ 0.80 was considered a large effect size difference.[112] For chi-squared analyses a percentage difference was also calculated from pre- to post-test.

3.5 Animal Care and Use and Ethics

All experimentation was in compliance with the University of Calgary Animal Care and Use Committee and in accordance with the Canadian Council of Animal Care, as well as the University of Calgary Conjoint Health Research Ethics Board.

Chapter Four: **Results**

4.1 Validity of eye-tracking technology

4.1.1 Concurrent Validity

I concurrently collected the verbal responses of participants using the think-aloud protocol with the eye-tracking recording data to establish criterion-related (concurrent) validity. In particular, the eye-tracking data showed concordant results with the think aloud protocol data in that experts spent significantly more time in the designated areas of interest (AOIs) than novices ($p < 0.017$) (Table 4.13) and identified significantly more of the key diagnostic features in their think-aloud responses than novices ($p < 0.017$) (Table 4.15). The AOIs selected *a priori* by two board certified clinical pathologists (Drs. Amy Warren and Catherine Wagg) was based on their clinical expertise and the published key features recognized for these five types of tumors.[110] As shown in Figure 4.2, experts had more efficient patterns of eye-movement with the eye-tracking when viewing the slide images than novices and had higher response levels of interpretation than novices from the think aloud data. In addition, experts took less time than novices to formulate a diagnosis with both the concurrent think-aloud protocol and without.

Supporting the construct validity of our data, and consistent with novice to expert studies in other visual fields,[6, 69] experts showed significantly higher diagnostic accuracy than novices ($p < 0.017$) (Table 4.14), took significantly less time than novices to come to a diagnosis ($p < 0.017$), and spent more percentage time in areas of diagnostic interest ($p < 0.017$) (Table 4.13). From the three group educational interventions study, the students' post-test results at the end of their respective intervention session had moved closer to the expert performance levels, such that there was no significant difference between experts and post-educational intervention novices for diagnostic accuracy ($p > 0.017$) (Table 4.40), no significant difference between post-test students

in Groups 1 (traditional) and 2 (basic visual) and experts for time to diagnosis ($p > 0.017$), and no significant difference between post-test students in Group 3 (extended visual) and experts for percentage time spent in the AOI ($p > 0.017$) (Table 4.39).

4.2 Novice and Expert Study

4.2.1 Descriptive statistics

Eleven student novices and six pathology experts participated voluntarily in the "novice and expert" study. There were 10 (91%) female and 1 (9%) male novices, and all 6 experts were female (100%). As stated above in the Methods section, the participants in the novice and expert groups were further divided into two groups where novice 1 group ($n = 6$) were instructed to use the think aloud protocol initially with slide images 1 to 5 followed by instructions not to use the protocol with the remaining images 6 to 10 (i.e., eye-tracking only). For novice 2 group ($n = 5$), the order of the two sets of images were reversed such that slide images 6 to 10 were introduced initially (i.e., eye-tracking only) followed by images 1 to 5 where the use of the think aloud protocol was introduced. This identical procedure for the sequencing of the slide images and use of the think aloud protocol was followed for the expert 1 ($n = 4$) and 2 ($n = 2$) groups.

Overall means and standard deviations, minimum and maximum scores, skewness and kurtosis for the continuous dependent variables time (seconds [s]) to diagnosis and percentage of time spent in AOI (%) for each of the novice and expert groups are provided in Table 4.1. Diagnostic accuracy, as a dependent dichotomous variable, was recorded for each slide (correct = 1, incorrect = 0) and total scores added for each score to calculate percent correct for individual participants as well as groups novice 1, novice 2, novice overall, expert 1, expert 2 and expert overall (Table 4.2) and for each image (Table 4.3). Mean and standard deviations for the time (s)

to diagnosis and percentage time spent in the AOIs for each slide (images 1-10) are in Table 4.3. There was a high positive skewness on the time to diagnosis for expert group 1 and experts overall.

Table 4.1 Minimum, maximum, mean (*SD*), skewness (*SE*) and kurtosis (*SE*) for continuous dependent variables time to diagnosis (s) and percentage time spent in the AOIs for groups novice 1, novice 2, novice overall, expert 1, expert 2 and expert overall.

Group	Viewed Slides	Variable	Min.	Max.	Mean (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)
Novice 1 (*n* = 6)	60	Time (s)	11.0	137.0	56.3 (30.9)	0.6 (0.3)	-0.2 (0.6)
		% time in AOI	7.0	81.0	41.9 (16.5)	-0.1 (0.3)	-0.3 (0.6)
Novice 2 (*n* = 5)	50	Time (s)	4.0	137.3	50.7 (26.5)	0.5 (0.3)	0.9 (0.7)
		% time in AOI	18.0	76.0	43.9 (13.1)	0.1 (0.3)	-0.3 (0.6)
Novice overall (*n* = 11)	110	Time (s)	4.0	137.3	53.8 (29.0)	0.6 (0.2)	0.2 (0.5)
		% time in AOI	7.0	81.0	42.8 (15.0)	-0.1 (0.2)	-0.2 (0.5)
Expert 1 (*n* = 4)	40	Time (s)*	0.1	99.2	18.1 (25.9)	2.2 (0.4)	4.1 (0.7)
		% time in AOI	10.4	100.0	60.3 (20.2)	0.1 (0.4)	0.0 (0.7)
Expert 2 (*n* = 2)	20	Time (s)*	1.1	43.0	13.6 (14.1)	1.2 (0.5)	-0.2 (1.0)
		% time in AOI	23.0	100.0	59.5 (18.9)	0.1 (0.5)	-0.1 (1.0)
Expert overall (*n* = 6)	60	Time (s)*	0.1	99.2	16.6 (2.9)	2.3 (0.3)	5.5 (0.6)
		% time in AOI	10.4	100.0	60.0 (19.6)	0.1 (0.3)	-0.1 (0.6)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.2 Diagnostic accuracy (% correct) for novice 1, novice 2, novice overall, expert 1, expert 2 and expert overall.

Group	*n*	Diagnostic accuracy (% correct)
Novice 1	60	11.7
Novice 2	50	12.0
Novice overall	110	11.8
Expert 1	40	97.5
Expert 2	20	85.0
Expert overall	60	93.3

Table 4.3 Mean and *SD* for the independent variables time to diagnosis (s) and percentage time spent in the AOIs and the diagnostic accuracy percentage correct for images 1-10 (n = 17).

	Image Number§	Time (s) *M* (*SD*)	% time in AOI *M* (*SD*)	Diagnostic accuracy % correct
Think aloud and eye-tracking	CH	60.7 (34.1)	50.0 (15.8)	35.3
	2 LSA	53.5 (30.1)	49.8 (10.4	52.9
	3 TVT	48.6 (32.9)	41.3 (11.8)	35.3
	MCT	45.7 (37.6)	59.0 (18.2)	58.8
	5 PCT	58.4 (31.2)	41.5 (9.1)	35.3
Eye-tracking only	CH	31.7 (26.9)	52.8 (24.8)	29.4
	7 PCT	29.1 (26.3)	47.5 (13.1)	35.3
	8 TVT	23.6 (21.2)	65.4 (13.0)	35.3
	MCT	29.6 (33.9)	48.9 (28.3)	52.3
	1 LSA	25.5 (22.0)	32.6 (13.8)	35.3

§CH= cutaneous histiocytoma; LSA = cutaneous lymphoma; TVT = transmissible venereal tumor; MCT = mast cell tumor; PCT = extramedullary plasmacytoma

4.2.2 Between group differences for novice 1 vs. novice 2 and expert 1 vs. expert 2

The order of the "think aloud" protocol slide images (1 to 5) and the "eye-tracking only" slide images (6 to 10) were reversed for the novices and experts (Table 3.1), with novice 1 and expert 1 instructed to use the think aloud protocol slides first, and novice 2 and expert 2 using the eye-tracking only slides first. To ascertain if there was a within group difference for novices and experts based on the order of the think aloud protocol, time to diagnosis (Table 4.4) and percentage time spent in the AOIs (Table 4.5) were compared using an independent samples *t*-test, and diagnostic accuracy compared using a chi-square test (Table 4.6).

There were no statistically significant differences between the mean performances of the novice 1 and novice 2, and expert 1 and expert 2 groups for time to diagnosis ($p > 0.017$), and percentage time spent in the AOIs ($p > 0.017$). In addition, no significant differences of percentage diagnoses correct (i.e., diagnostic accuracy) were found between novice 1 and novice 2 and expert 1 and expert 2 ($p > 0.017$). Thus, the sequencing of the think aloud slides (images 1

to 5) and eye-tracking only slides (images 6 to 10) had no impact on the performance of the novices and experts groups in the study.

Table 4.4 Independent samples *t*-test results (*n*, *t*-value, *df*, and *p*) and effect size difference (*d*) for time to diagnosis for novice 1 vs. novice 2 and expert 1 vs. expert 2.

Group	Slide *n*	Time to diagnosis (s) *M* (*SD*)	*t*-value	*df*	*p*	Effect size Cohen's *d* (95% CI)
Novice 1 (*n* = 6)	60	56.3 (30.9)	0.995	108	0.322	0.19 (-0.19-0.56)
Novice 2 (*n* = 5)	50	50.7 (26.5)				
Expert 1 (*n* = 4)	40	18.1 (25.9)	0.714	58	0.478	0.20 (-0.34 – 0.73)
Expert 2 (*n* = 2)	20	13.6 (14.1)				

Table 4.5 Independent samples *t*-test results (*n*, *t*-value, *df*, and *p*) and effect size difference (*d*) for percentage time spent in the AOIs for novice 1 vs. novice 2 and expert 1 vs. expert 2.

Group	Slide *n*	Percentage (%) time in AOI *M* (*SD*)	*t*-value	*df*	*p*	Effect size Cohen's *d* ǂ (95% CI)
Novice 1 (*n* = 6)	60	41.9 (16.5)	-0.696	108	0.488	-0.13 (-0.51 - 0.24)
Novice 2 (*n* = 5)	50	43.9 (13.1)				
Expert 1 (*n* = 4)	40	60.3 (20.2)	0.152	68	0.879	0.04 (-0.50 – 0.58)
Expert 2 (*n* = 2)	20	59.5 (18.9)				

Table 4.6 Chi-square between group differences for diagnostic accuracy (percent correct) for novice 1 vs. novice 2 and expert 1 vs. expert 2

Group	*n*	Diagnostic accuracy (% correct)	Chi-square value	*df*	*p*
Novice 1 (*n* = 6)	60	11.6	0.003	1	0.957
Novice 2 (*n* = 5)	50	12.0			
Expert 1 (*n* = 4)	40	97.5	3.348	1	0.067
Expert 2 (*n* = 2)	20	85.0			

4.2.3 Differences in "talk aloud" slides (TA) (images 1-5) and "eye-tracking only" (ET) slides (images 6-10)

4.2.3.1 Quantitative differences between TA and ET slides

The time to diagnosis, percentage time spent in the AOIs and diagnostic accuracy were combined for novice and expert groups for the TA slides (images 1 to 5) and the ET slides (images 6 to 10), and were compared using an independent samples *t*-test (time and percentage time in the AOIs) (Table 4.7) and a chi-square test (Table 4.8). Similarly, the time to diagnosis, percentage time spent in the AOIs and diagnostic accuracy were compared for the TA and ET slides for the novice and expert groups separately using a paired-sample *t*-test (Table 4.9) and chi-square test (Table 4.10).

For all participants combined, there was a significant difference between the means of TA (images 1 to 5) (M = 53.4s, SD = 33.0s) for time to diagnosis compared to the mean of ET (images 6-10) (M = 27.9s, SD = 26.0s), $t(159.4) = 5.60$, $p < 0.017$. For percentage time spent in the AOI and diagnostic accuracy, there was no significant difference between the TA (images 1-5) and ET (images 6-10) slides, ($p > 0.017$).

Similarly, for each of the novice and expert groups, there was a significant difference between TA and ET slides for time to diagnosis for the novices (M = 67.3s, SD = 26.8s; M = 40.2s, SD = 24.6s, respectively) $t(108) = 5.531$, $p < 0.017$, and for the experts (M = 27.8s, SD = 27.7s; M = 5.3s, SD = 3.5s, respectively) $t(30) = 4.413$, $p < 0.017$. No significant difference was found for the novice and expert groups between TA and ET slides for percentage time spent in the AOIs ($p > 0.017$) and diagnostic accuracy ($p > 0.017$).

Table 4.7 Independent samples *t*-test comparing the mean time to diagnosis (s) and percentage time in the AOIs (%) for the aggregated novice and expert results for the "think aloud" (TA) slides 1 to 5 compared to the aggregated novice and expert results for the "eye-tracking only" (ET) slides 6 to 10.

Parameter	**TA (images 1-5) (*n* = 85) *M* (*SD*)**	**ET (Images 6-10) (*n* = 85) *M* (*SD*)**	***t*-value**	***df***	***p***	**Effect size Cohen's *d* ‡ (95% CI)**
Time (s) §	53.4 (33.0)	27.9 (26.0)	5.600	159.4	0.001†	0.86 (0.54 – 1.17)
% time in AOI§	48.3 (14.7)	49.4 (21.9)	-0.395	147.0	0.693	-0.11 (-0.41 – 0.19)

†$p < 0.001$. § Levene's Test for Equality of Variances not equal, therefore equal variances not assumed. ‡ Cohen's $d > 0.80$ large effect size difference.

Table 4.8 Chi-square test comparing the diagnostic accuracy (percentage correct) for the aggregated novice and expert results for the "think-aloud" (TA) slides 1 to 5 compared to the aggregated novice and expert results for the "eye-tracking only" (ET) slides 6 to 10.

Group	***n***	**Diagnostic accuracy (% correct)**	**Chi-square value**	***df***	***p***	**Percentage difference**
TA (Images 1-5) (*n* = 85)	85	43.5	0.619	1	0.435	5.9
ET (Images 6-10)	85	37.6				

Table 4.9 Paired sample *t*-test comparing the mean time to diagnosis (s) and percentage time spent in the AOIs (%) for the "think-aloud" (TA) slides 1 to 5 compared to the aggregated novice and expert results for the "eye-tracking only" (ET) slides 6 to 10 for the novice and expert groups separately.

	n	**Parameter**	**TA (images 1-5) *M* (*SD*)**	**ET (Images 6-10) *M* (*SD*)**	***t*-value**	***df***	***p***	**Effect size Cohen's *d* ‡ (95% CI)**
Novices (n = 11)	110	Time (s)	67.3 (26.8)	40.2 (24.6)	5.531	108	<0.001†	1.05 (0.66 – 1.45)
		% time in AOI§	44.4 (11.6)	41.2 (17.7)	-1.143	93.1	0.256	-0.22 (-0.59 – 0.16)
Experts (n = 6)	60	Time (s) §	27.8 (27.7)	5.3 (3.5)	4.413	30.0	<0.001†	1.14 (0.59 – 1.69)
		% time in AOI	55.4 (17.2)	64.6 (21.0)	-1.855	58	0.069	-0.48 (-0.99 – 0.03)

†$p < 0.001$. § Levene's Test for Equality of Variances not equal therefore equal variances not assumed. ‡ Cohen's $d > 0.80$ large effect size difference.

Table 4.10 Chi-square test comparing the diagnostic accuracy (percentage correct) for the "think-aloud" (TA) slides 1 to 5 compared to the "eye-tracking only" (ET) slides 6-10 for each of the novice and expert groups separately.

Group	**TA/ ET**	***n***	**Diagnostic accuracy (% correct)**	**Chi-square value**	***df***	***p***	**Percentage difference**
Novice (n = 11)	TA (Images 1-5)	55	16.4	2.181	1	0.140	9.1
	ET (Images 6-10)	55	7.3				
Expert (n = 6)	TA (Images 1-5)	30	93.3	0.000	1	1.000	0
	ET (Images 6-10)	30	93.3				

4.2.3.2 Qualitative differences between TA and ET slides

Patterns of eye-movement from TA (images 1-5) and ET (images 6-10) slides were compared for both novices and experts. Although novices took a longer time to view the TA images than ET images, the patterns of eye-movement were so complex in both TA and ET images that no visually discernable difference was noted between the two groups. Experts, however, had simpler and more directed patterns of eye-movement (see section 4.2.6.4) and there were qualitative differences shown between the TA and ET slides. The eye-movement patterns for experts in the TA slides were more complex, with more points of fixation on the AOIs, though a clear pattern of movement was still discernable. In ET slides, eye-movement patterns were simple and focused relatively longer on the AOIs, requiring much less time before a diagnosis was provided by the expert (Figure 4.1). The exception was for expert 4, who in the think aloud slides came straight to a diagnosis without verbalizing, such that her eye-movement patterns for both TA and ET slides were similar (Figure 4.1).

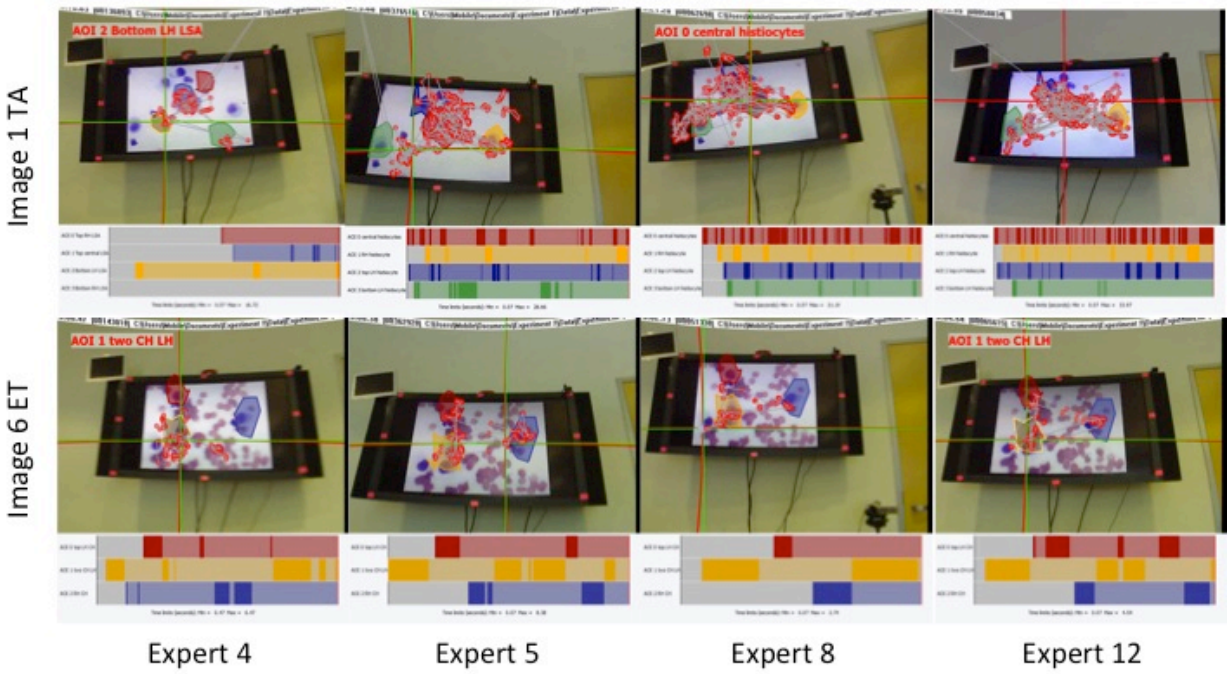


Figure 4.1 Eye-movement patterns for experts for cutaneous histiocytoma in the think aloud (TA) section (image 1) and the eye-tracking only (ET) section (image 6).

4.2.4 Differences between novices and experts for individual diagnoses

As there was no significant difference for the time (s) to diagnosis, percentage time spent in the AOIs (%) and diagnostic accuracy (% correct) between novice 1 and novice 2 groups, and expert 1 and expert 2 groups, the novice and expert results on the dependent variables were combined together for novice vs. expert comparisons for individual slides initially, and later as aggregated data.

As this was a preliminary study, the time to diagnosis and percentage time spent in the AOIs for the novices and experts were compared for each slide using an independent samples *t*-test (Table 4.11). The diagnostic accuracy for the novices and experts for each slide were compared using a chi-square test (Table 4.12).

For image 1 (cutaneous histiocytoma), there was no significant difference between the means of novices and experts for time to diagnosis and percentage time spent in the AOIs, ($p >$ 0.017). There was a significant difference found between novices and experts for percentage diagnoses (0% vs. 100%, respectively) correct (i.e., diagnostic accuracy), χ^2 (1, N = 17) = 17.00, $p < 0.017$.

For image 2 (cutaneous lymphoma), novices took significantly more time (M = 68.1s, SD = 23.3s) than experts (M = 26.6s, SD = 21.6s), $t(15) = 3.604$, $p < 0.017$, to reach a diagnosis. There was no significant difference between the means of novices and experts for the percentage time spent in the AOIs, ($p > 0.017$). There was a significant difference between novices and experts for percentage diagnoses correct (27.3% vs. 100%, respectively), χ^2 (1, N = 17) = 8.242, $p < 0.017$.

For image 3 (transmissible venereal tumor), there was no significant difference between the novices and experts for the time to reach a diagnosis and percentage time spent in the AOIs, ($p > 0.017$). There was a significant difference between novices and experts for percentage diagnoses correct (0% vs. 100%, respectively) χ^2 (1, N = 17) = 17.000, $p < 0.017$.

For image 4 (mast cell tumor), novices took on average significantly more time (M = 65.8s, SD = 30.3s) than experts (M = 8.8s, SD = 12.1s) [$t(15) = 4.367$, $p < 0.017$] to reach a diagnosis and novices spent on average less percentage time in the AOIs (M = 49.8%, SD = 10.8%) compared to experts (M = 76.0%, SD = 17.2%) [$t(15) = -3.896$, $p < 0.017$]. There was a significant difference between novices and experts for percentage diagnoses correct (36.4% vs. 100%, respectively) χ^2 (1, N = 17) = 6.491, $p < 0.017$.

For image 5 (extramedullary plasmacytoma), there was no significant difference between the means of novices and experts for the time to diagnosis, the percentage time spent in the AOIs, and the diagnostic accuracy ($p > 0.017$).

For image 6 (cutaneous histiocytoma), novices took on average significantly more time ($M = 45.6$s, $SD = 23.4$s) than experts ($M = 6.1$s, $SD = 2.1$s) [$t(15) = 5.573$, $p < 0.017$] to reach a diagnosis and novices spent less percentage time in the AOIs ($M = 41.0\%$, $SD = 22.6\%$) compared to experts ($M = 74.5\%$, $SD = 8.4\%$) [$t(15) = -4.405$, $p < 0.017$]. There was a significant difference between novices and experts for percentage diagnoses correct (0% vs. 83.3%, respectively) $\chi^2 (1, N = 17) = 12.986$, $p < 0.017$.

For image 7 (extramedullary plasmacytoma), novices took significantly more time ($M = 41.7$s, $SD = 24.4$s) than experts ($M = 5.7$s, $SD = 2.9$s) [$t(15) = 3.551$, $p < 0.017$] to reach a diagnosis and novices spent less percentage time in the AOIs ($M = 40.3\%$, $SD = 8.0\%$) compared to experts ($M = 60.8\%$, $SD = 9.8\%$) [$t(15) = -4.671$, $p < 0.017$]. There was a significant difference between novices and experts for percentage diagnoses correct (0% vs. 100%, respectively) $\chi^2 (1, N = 17) = 17.000$, $p < 0.017$.

For image 8 (transmissible venereal tumor), novices took significantly more time ($M = 33.1$s, $SD = 20.8$s) than experts ($M = 6.1$s, $SD = 2.5$s) to reach a diagnosis, $t(10.5) = 4.245$, $p < 0.017$. There was no significant difference between novices and experts for the percentage time spent in the AOIs, ($p > 0.017$). There was a significant difference between novices and experts for percentage diagnoses correct (0% vs. 100%, respectively) $\chi^2 (1, N = 17) = 17.000$, $p < 0.017$.

For image 9 (mast cell tumor), novices took significantly more time ($M = 44.9$s, $SD = 3.33$s) than experts ($M = 1.7$s, $SD = 1.0$s) [$t(15) = 4.299$, $p < 0.017$] to reach a diagnosis. There was no significant difference between novices and experts for the percentage time spent in the

AOIs, $p < 0.017$. There was a significant difference between novices and experts for percentage diagnoses correct (27.3% vs. 100%, respectively) χ^2 (1, $N = 17$) = 8.242, $p < 0.017$.

For image 10 (cutaneous lymphoma), novices took significantly more time (M = 35.6s, SD = 21.1s) than experts (M = 7.0s, SD = 5.6s), $t(15) = 3.216$, $p < 0.017$ to reach a diagnosis. There was no significant difference between novices and experts for the percentage time spent in the AOIs, $p < 0.017$. There was a significant difference between novices and experts for percentage diagnoses correct (18.2% vs. 100%, respectively), χ^2 (1, $N = 17$) = 8.242, $p < 0.017$.

Table 4.11 Independent samples *t*-test comparing the mean time to diagnosis (s) and percentage time in the AOIs (%) for the aggregated novice and expert results for images 1-10.

Slide¶	Parameter	Novice (n = 11) M (SD)	Expert (n = 6) M (SD)	*t*-value	*df*	*p*	Effect size Cohen's *d* ‡ (95% CI)
1 CH	Time (s)	71.9 (33.1)	40.0 (26.9)	2.019	15	0.062	1.02 (0.03-2.07)
	% time in AOI	48.4 (13.0)	52.8 (21.1)	-0.539	15	0.598	-0.27 (-1.27 – 0.73)
2 LSA	Time (s)	68.1 (23.3)	26.6 (21.6)	3.604	15	0.003*	1.83 (0.66 -3.00)
	% time in AOI	46.4 (9.2)	56.0 (10.1)	-1.984	15	0.066	-1.01 (-2.06 – 0.04)
3 TVT	Time (s)	60.4 (25.0)	26.9 (36.6)	2.243	15	0.040	1.14 (0.07 – 2.21)
	% time in AOI	38.3 (12.4)	46.7 (9.2)	-1.440	15	0.170	-0.73 (-1.76 – 0.29)
4 MCT	Time (s)	65.8 (30.3)	8.8 (12.1)	4.367	15	0.001†	2.22 (0.97 – 3.46)
	% time in AOI	49.8 (10.8)	76.0 (17.2)	-3.896	15	0.001†	-1.98 (-3.17 - -0.78)
5 PCT	Time (s)	70.2 (24.5)	36.8 (32.5)	2.400	15	0.030	1.22 (0.14 – 2.29)
	% time in AOI	39.2 (8.9)	45.7 (8.7)	-1.430	15	0.173	-0.73 (-1.75 – 0.30)
6 CH	Time (s)§	45.6 (23.4)	6.1 (2.1)	5.573	10.3	<0.001†	2.83 (1.45 – 4.21)
	% time in AOI§	41.0 (22.6)	74.5 (8.4)	-4.405	13.9	0.001†	-2.24 (-3.48 - -0.99)
7 PCT	Time (s)	41.7 (24.4)	5.7 (2.9)	3.551	15	0.003*	1.80 (0.64 – 2.97)
	% time in AOI	40.3 (8.0)	60.8 (9.8)	-4.671	15	<0.001†	-2.37 (-3.65 - -1.10)
8 TVT	Time (s) §	33.1 (20.8)	6.1 (2.5)	4.245	10.5	0.002*	2.15 (0.92 – 3.38)
	% time in AOI	61.3 (11.5)	72.9 (13.1)	-1.895	15	0.078	-0.96 (-2.01 – 0.08)
9 MCT	Time (s) §	44.9 (33.3)	1.7 (1.0)	4.299	10.0	0.002*	2.18 (0.95 – 3.42)
	% time in AOI§	35.5 (14.7)	73.3 (31.9)	-2.750	6.2	0.032	-1.40 (-2.50 - -0.30)
10 LSA	Time (s)	35.6 (21.1)	7.0 (5.6)	3.216	15	0.006*	1.63 (0.50 – 2.77)
	% time in AOI	27.7 (9.8)	41.7 (16.1)	-2.238	15	0.041	-1.14 (-2.20 - - 0.07)

* $p < 0.01$, † $p < 0.001$. ¶CH= cutaneous histiocytoma; LSA = cutaneous lymphoma; TVT = transmissible venereal tumor; MCT = mast cell tumor; PCT = extramedullary plasmacytoma. § Levene's Test for Equality of Variances not equal therefore equal variances not assumed. ‡ Effect size difference (Cohen's *d):* small = 0.30 to 0.49, medium = 0.50 to 0.79, > 0.80 large.

Table 4.12 Chi-square test for diagnostic accuracy (percent correct) of the aggregated novice and expert results for images 1-10.

Image ¶	Novice % correct (n = 11)	Expert % correct frequency (n = 6)	Chi-square value	*df*	*p*	Percentage difference (%)
1 CH	0	100.0	17.000	1	<0.001†	100.0
2 LSA	27.3	100.0	8.242	1	0.004*	72.7
3 TVT	0	100.0	17.000	1	<0.001†	100.0
4 MCT	36.4	100.0	6.491	1	0.011*	63.6
5 PCT	18.2	66.7	3.996	1	0.046	48.5
6 CH	0	83.3	12.986	1	<0.001†	83.3
7 PCT	0	100.0	17.000	1	<0.001†	100.0
8 TVT	0	100.0	17.000	1	<0.001†	100.0
9 MCT	27.3	100.0	8.242	1	0.004*	72.7
10 LSA	18.2	83.3	9.370	1	0.002*	65.1

*$p < 0.017$, †$p < 0.001$. ¶CH= cutaneous histiocytoma; LSA = cutaneous lymphoma; TVT = transmissible venereal tumor; MCT = mast cell tumor; PCT = extramedullary plasmacytoma.

4.2.5 Differences between novices and experts for aggregated data

Time (s) to diagnosis and percentage time in the AOIs (%) results were combined for all slides (images 1 to10), for all novices (novice 1 and novice 2 groups) and for all experts (expert 1 and expert 2 groups). A student's *t*-test for independent samples was performed to analyze differences between novices and experts for the mean time to diagnosis (s) and percentage time spent in the AOIs (%) (Table 4.13). Diagnostic accuracy (% of correct diagnoses) results were combined for all slides (images 1 to 10), for all novices and for all experts. A chi-square test was used to compare differences between novice and expert groups (Table 4.14).

For time (s) to diagnosis, novices (M = 53.8s, SD = 16.6s) took significantly more time to reach a diagnosis compared to experts (M = 16.6s, SD = 22.6s), $t(147.8) = 9.246$, $p < 0.017$. Novices (M = 42.8%, SD = 15.0%) spent less percentage time in the AOIs compared to experts

(M = 60.0%, SD = 19.6%), $t(97.4) = -5.292$, $p < 0.017$. There was a significant difference between novices and experts for diagnostic accuracy (11.8% vs. 93.3%, respectively) χ^2 (1, N = 170) = 106.979, $p < 0.017$.

Table 4.13 Independent samples *t*-test comparing the mean time to diagnosis (s) and percentage time in the AOIs (%) for the aggregated novice (n =11) and expert (n = 6) results for images 1 to 10.

Parameter	**Novice aggregated (n = 110) M (SD)**	**Expert aggregated (n = 60) M (SD)**	***t*-value**	***df***	***p***	**Effect size Cohen's *d* ‡ (95% CI)**
Time (s) §	53.8 (16.6)	16.6 (22.6)	9.246	147.8	<0.001†	1.48 (1.13 – 1.84)
% time in AOI§	42.8 (15.0)	60.0 (19.6)	-5.929	97.4	<0.001†	-0.95 (-1.28 - -0.62)

†$p < 0.001$. § Levene's Test for Equality of Variances not equal therefore equal variances not assumed. ‡ Cohen's d > 0.80 large effect size difference.

Table 4.14 Chi-square test for between group differences for diagnostic accuracy (percent correct) for novice aggregated vs. expert aggregated.

Group	***N***	**Diagnostic accuracy (% correct)**	**Pearson chi-square value**	***df***	***p***	**Percentage difference (%)**
Novice overall	110	11.8	106.979	1	<0.001†	81.5
Expert overall	60	93.3				

†$p < 0.001$.

4.2.6 Qualitative novice-expert data

4.2.6.1 Key diagnostic features

For each slide (images 1 to 10), between 4 and 9 diagnostic features were identified by the experts in the concurrent (i.e., images 1 to 5, where participants were asked to talk aloud while eye-tracking was monitored) and retrospective think aloud protocols used on all of the slide images. The frequency of the features identified by the participants was recorded for both the experts and the novices. Of these, the two to four features most frequently identified by experts (> 50.0% of experts) that were pertinent to differentiating the diagnosis were identified as the "key diagnostic features". The terms "round cell" or "discrete/ individual cells" were excluded as these are features common to the whole group of round cell tumor diagnoses used in the study. The frequency that experts mentioned the key diagnostic features was compared to the frequency that novices mentioned these same features using a chi-square test (Table 4.15).

For the cutaneous histiocytoma (image 1), the key diagnostic features identified were a moderate amount of cytoplasm, infiltrating small lymphocytes, moderate cell size and monomorphic cell population. There was a significant difference between novices (18.2%) and experts (100.0%) for observing a moderate amount of cytoplasm χ^2 (df = 1, N = 17) = 10.432, p < 0.017, between novice (100.0%) and experts (18.2%) for observing infiltrating small lymphocytes χ^2 (df = 1, N = 17) = 10.432, p < 0.017), but no significant difference between novices and experts for observing moderate cell size (p > 0.017) and a monomorphic cell population (p > 0.017). For the second image of the cutaneous histiocytoma (image 6), the key diagnostic features identified were a moderate amount of cytoplasm, clear to pale-blue staining cytoplasm and a monomorphic cell population. There was a significant difference between novices (18.2%) and experts (83.3%) for observing a moderate amount of cytoplasm χ^2 (df = 1, N

= 17) = 6.804, $p < 0.017$, between novices (0%) and experts (66.7%) for observing clear to pale-blue staining cytoplasm χ^2 ($df = 1$, $N = 17$) = 9.590, $p < 0.017$ and between novices (0%) and experts (50.0%) for observing a monomorphic cell population χ^2 ($df = 1$, $N = 17$) = 6.679, $p < 0.017$.

For the cutaneous lymphoma (image 2), the key diagnostic features identified were a monomorphic cell population, large lymphocytes and a mitotic figure. There was a significant difference between novices (9.1%) and experts (83.3%) for observing a monomorphic cell population χ^2 ($df = 1$, $N = 17$) = 9.370, $p < 0.017$, and between novices (0%) and experts (83.3%) for observing a mitotic figure χ^2 ($df = 1$, $N = 17$) = 12.986, $p < 0.017$. There was no significant difference between novices and experts for observing large lymphocytes. For the second image of the cutaneous lymphoma (image 10), the key diagnostic features identified were the presence of large lymphocytes, a small amount of cytoplasm, a fine chromatin pattern and the presence of lymphoglandular bodies. There was a significant difference between the novices (9.1%) and experts (83.3%) for observing the presence of large lymphocytes χ^2 ($df = 1$, $N = 17$) = 9.370, $p < 0.017$ and between novices (0%) and experts (50.0%) for observing the presence of lymphoglandular bodies χ^2 ($df = 1$, $N = 17$) = 6.679, $p < 0.017$, but no significant difference between novices and experts for observing a small amount of cytoplasm ($p > 0.017$) and a fine chromatin pattern ($p > 0.017$).

For the transmissible venereal tumor (image 3), the key diagnostic features identified were cytoplasmic vacuolation, a moderate amount of cytoplasm, a ropey chromatin pattern and a moderate degree of anisokaryosis. There was a significant difference between novices (0%) an experts (66.7%) for observing a moderate amount of cytoplasm χ^2 ($df = 1$, $N = 17$) = 9.590, $p < 0.017$ and between novices (0%) and experts (50.0%) for observing a ropey chromatin pattern χ^2

($df = 1$, $N = 17$) = 6.679, $p < 0.017$, but no significant difference between novices and experts for observing cytoplasmic vacuolation ($p > 0.017$) and a moderate degree of anisokaryosis ($p > 0.017$). For the second image of the transmissible venereal tumor (image 8), the key diagnostic features identified were cytoplasmic vacuolation, a ropey chromatin pattern and a moderate amount of cytoplasm. There was a significant difference between novices (0%) and experts (50.0%) for observing a moderate amount of cytoplasm χ^2 ($df = 1$, $N = 17$) = 6.679, $p < 0.017$, but no significant difference between novices and experts for observing ropey chromatin pattern or cytoplasmic vacuolation ($p > 0.017$).

For the mast cell tumor (image 4), the key diagnostic features identified were purple cytoplasmic granules and infiltrating lymphocytes. There was no significant difference between novices and experts for observing purple cytoplasmic granules ($p > 0.017$) and infiltrating lymphocytes ($p > 0.017$). For the second mast cell tumor (image 9), the key diagnostic features identified were purple cytoplasmic granules and the presence of fibroblasts. There was a significant difference between novices (18.2%) and experts (83.3%) for observing the presence of fibroblasts χ^2 ($df = 1$, $N = 17$) = 6.804, $p < 0.017$, but no significant difference between novices and experts for observing purple cytoplasmic granules ($p > 0.017$).

For the extramedullary plasmacytoma (image 5), the key diagnostic features identified were medium-blue cytoplasm, the presence of a peri-nuclear clear zone, eccentrically placed nuclei and a moderate amount of cytoplasm. There was a significant difference between novices (0%) and experts (50.0%) for observing a peri-nuclear clear zone χ^2 ($df = 1$, $N = 17$) = 6.679, $p < 0.017$, but no significant difference between novices and experts for observing medium-blue cytoplasm, eccentrically placed nuclei, and a moderate amount of cytoplasm ($p > 0.017$). For the second extramedullary plasmacytoma (image 7), the key diagnostic features identified were peri-

nuclear clear zone, deep blue cytoplasm and eccentrically placed nucleus. There was a significant difference between novices (0%) and experts (100.0%) for observing a peri-nuclear clear zone χ^2 (df = 1, N = 17) = 17.000, $p < 0.017$ and between novices (18.2%) and experts (100.0%) for observing deep blue cytoplasm χ^2 (df = 1, N = 17) = 10.432, $p < 0.017$. No significant difference was noted between novices and experts for observing eccentrically placed nuclei, ($p > 0.017$).

Table 4.15 Chi-square test comparing the frequency that novices and experts stated the key diagnostic features as identified by experts for each of slides 1-10 in think-aloud protocols.

Image ¶	Key diagnostic feature	Novice frequency (*n* = 11) (%)	Expert frequency (*n* = 6) (%)	Pearson chi-square value	*df*	*p*	Percent diff. (%)
1 CH	Moderate amount cytoplasm	18.2	100.0	10.432	1	0.001†	81.8
	Infiltrating small lymphocytes	18.2	100.0	10.432	1	0.001†	81.8
	Moderate cell size	27.3	50.0	0.878	1	0.349	22.7
	Monomorphic	9.1	50.0	3.611	1	0.057	40.9
2 LSA	Monomorphic	9.1	83.3	9.370	1	0.002*	74.2
	Large lymphocytes	27.3	83.3	4.898	1	0.027	56.0
	Mitotic figure	0	83.3	12.986	1	<0.001†	
3 TVT	Cytoplasmic vacuoles§	100	100.0	-	-	-	0
	Moderate amount of cytoplasm	0	66.7	9.590	1	0.002*	66.7
	Ropey chromatin pattern	0	50.0	6.679	1	0.010*	50.0
	Anisokaryosis	45.5	50.0	0.298	1	0.585	4.5
4 MCT	Purple cytoplasmic granules	72.7	100.0	1.987	1	0.159	27.8
	Infiltrating lymphocytes	45.5	66.7	0.701	1	0.402	21.2
5 PCT	Medium-blue cytoplasm	18.2	66.7	3.996	1	0.046	48.5
	Peri-nuclear clear zone	0	50.0	6.679	1	0.010*	50.0
	Eccentric nuclei	36.4	50.0	0.878	1	0.349	13.6
	Moderate amount of cytoplasm	18.2	50.0	1.571	1	0.210	31.8
6 CH	Moderate amount of cytoplasm	18.2	83.3	6.804	1	0.009*	65.1
	Clear to pale blue cytoplasm	0	66.7	9.590	1	0.002*	66.7
	Monomorphic	0	50.0	6.679	1	0.010*	50.0
7 PCT	Perinuclear clear zone	0	100.0	17.000	1	<0.001†	100.0
	Deep blue cytoplasm	18.2	100.0	10.432	1	0.001†	81.8
	Eccentric nuclei	18.2	66.7	3.996	1	0.046	48.5
8 TVT	Cytoplasmic vacuoles	63.6	100.0	2.853	1	0.091	36.4

	Ropey chromatin pattern	18.2	66.7	3.996	1	0.046	48.5
	Moderate amount of cytoplasm	0	50.0	6.679	1	0.010*	50.0
9 MCT	Purple cytoplasmic granules	63.6	100.0	2.853	1	0.090	36.4
	Fibroblasts	18.2	83.3	6.804	1	0.009*	65.1
10 LSA	Large lymphocytes	9.1	83.3	9.370	1	0.002*	74.2
	Small amount of cytoplasm	18.2	50.0	1.893	1	0.169	31.8
	Fine chromatin pattern	9.1	50.0	3.611	1	0.057	40.9
	Lymphoglandular bodies	0	50.0	6.679	1	0.010*	50.0

*$p < 0.017$,[†]$p < 0.001$. CH= cutaneous histiocytoma; LSA = cutaneous lymphoma; TVT = transmissible venereal tumor; MCT = mast cell tumor; PCT = extramedullary plasmacytoma. § Chi-squared analysis was not performed for this diagnosis as the novice and expert % are identical.

4.2.6.2 Diagnostic features from qualitative data compared to AOI

Areas of interest (AOI) were selected *a priori* by two board certified pathologists based on diagnostic experience and published diagnostic features.[110] Cells were selected for their "representativeness" of diagnostic features such that the cells when viewed alone were most likely to lead a viewer to the cytologic diagnosis. The key diagnostic features elicited from the think aloud protocols for most slides corresponded to the selected AOIs, with the exception of infiltrating lymphocytes in images 1 and 4 and fibroblasts in image 9. These cells were not tumor cells, however, their presence (in image 1 and 9) were supportive of the diagnosis.

4.2.6.3 Qualitative differences in novice and expert reasoning and visualization based on think-aloud data

In addition to identifying diagnostic features, there were several differences between the novices and expert groups that were elucidated from the concurrent think-aloud data (i.e. images 1-5): 1) there was a difference in the time to diagnosis between candidates who knew the

diagnosis and were correct, were unsure about the diagnosis, or had no idea about the diagnosis; 2) experts used multiple key diagnostic features to formulate a diagnosis, compared to novices who used very few; 3) novices observed more non-diagnostic or irrelevant features; and 4) experts used more accurate and interpretive terminology compared to novices.

When a diagnosis was known and correct (experts), the time to diagnosis was typically short. Similarly, if a candidate had no idea bout the diagnosis (novices), they also tended to state this more quickly. However, when novices or experts were unsure about a diagnosis, the time they spent viewing the image was longer.

Experts, generally, were able to verbalize more diagnostic features than novices. Experts typically used more than four diagnostic features to support their diagnosis, and used these features to differentiate their differential diagnoses "the presence of a peri-nuclear clear zone and the color of the cytoplasm (medium-blue) makes me think that this is more likely a plasmacytoma than a histiocytoma" (expert 24, image 5). In comparison, novices had a tendency to make decisions based on only one visual feature "I remember something about (purple) granules and mast cells so I guess a mast cell tumor?" (novice 10, image 4). In addition, when asked *"Based on your clinical diagnosis for this slide (provide participant with their response), how did you arrive at this decision"*, novices more frequently had a difficult time justifying their diagnosis in comparison to experts, citing typically only one diagnostic feature.

Novices more frequently observed non-diagnostic or irrelevant features. For example, on almost every slide, several of the novices mentioned the presence of erythrocytes (red blood cells) and some would erroneously interpret their presences as "evidence of haemorrhage" (novice 30, image 4) or "this must be a blood sample" (novice 19, image 2). Experts never mentioned erythrocytes, as erythrocytes in all these cases were evidence of blood contamination

in the sample collection and diagnostically irrelevant. Novices also more frequently commented on staining anomalies "I don't like the purple looking nuclei, I think that must mean there is something bad going on" (novice 17, image 3) whereas experts never mentioned the staining differences. This is supported by the eye-tracking data that experts spent a significantly higher percentage of time in the AOIs compared to novices.

When observing diagnostic features, experts were more frequently able to assign the correct terminology "mitotic figure" (expert 4, image 4) compared to novices who often used more basic descriptors of color and shape "there are two black dots in that round pink cell that I think means it is dividing" (novice 30, image 4). Experts also made statements indicating a higher level of interpretation "mitotic figure- this is a high grade mast cell tumor" (expert 4, image 4); "the cells are lymphoblasts and based on the presence of lymphoglandular bodies I would bet this is a B-cell lymphoma" (expert 4, image 2). Novices rarely made interpretations in their descriptions "the cells are large and haven't got much of an envelope (cytoplasm) around them" (novice 13, image 2) describing the same lymphoblasts.

Experts very quickly stated a diagnosis, or a few narrow differential diagnoses, then cited visual evidence in the image to support their diagnosis or to discriminate between their differential diagnoses. In image 2 and 3, experts 4 and 8 came straight to stating a diagnosis in the concurrent talk aloud section and in image 4, experts 4, 8, 24 and 25 came straight to a diagnosis with no talking. This is supported by the significantly shorter time it took experts to diagnose the images on the eye tracking. Novices tended to be analytic in their approach to the slide stating findings, interpreting their significance, and later collating the information together to come up with a diagnosis. At times, however, novices would attempt to recognize patterns, "This really reminds me of something. It is very familiar. I think the uniform nuclei means they

are hepatocytes?" (they were plasma cells) (novice 17, image 5) "I am hoping for some inspiration to trigger a memory of what I am looking at" (novice 30, image 1).

4.2.6.4 Eye-movement patterns

Experts generally had very efficient patterns of eye-movement, quickly identifying and efficiently moving between the diagnostic features (AOIs) on the slide, with minimal attention to the non-diagnostic areas. Novices on the contrary had erratic and complex patterns of eye-movement (Figure 4.2).

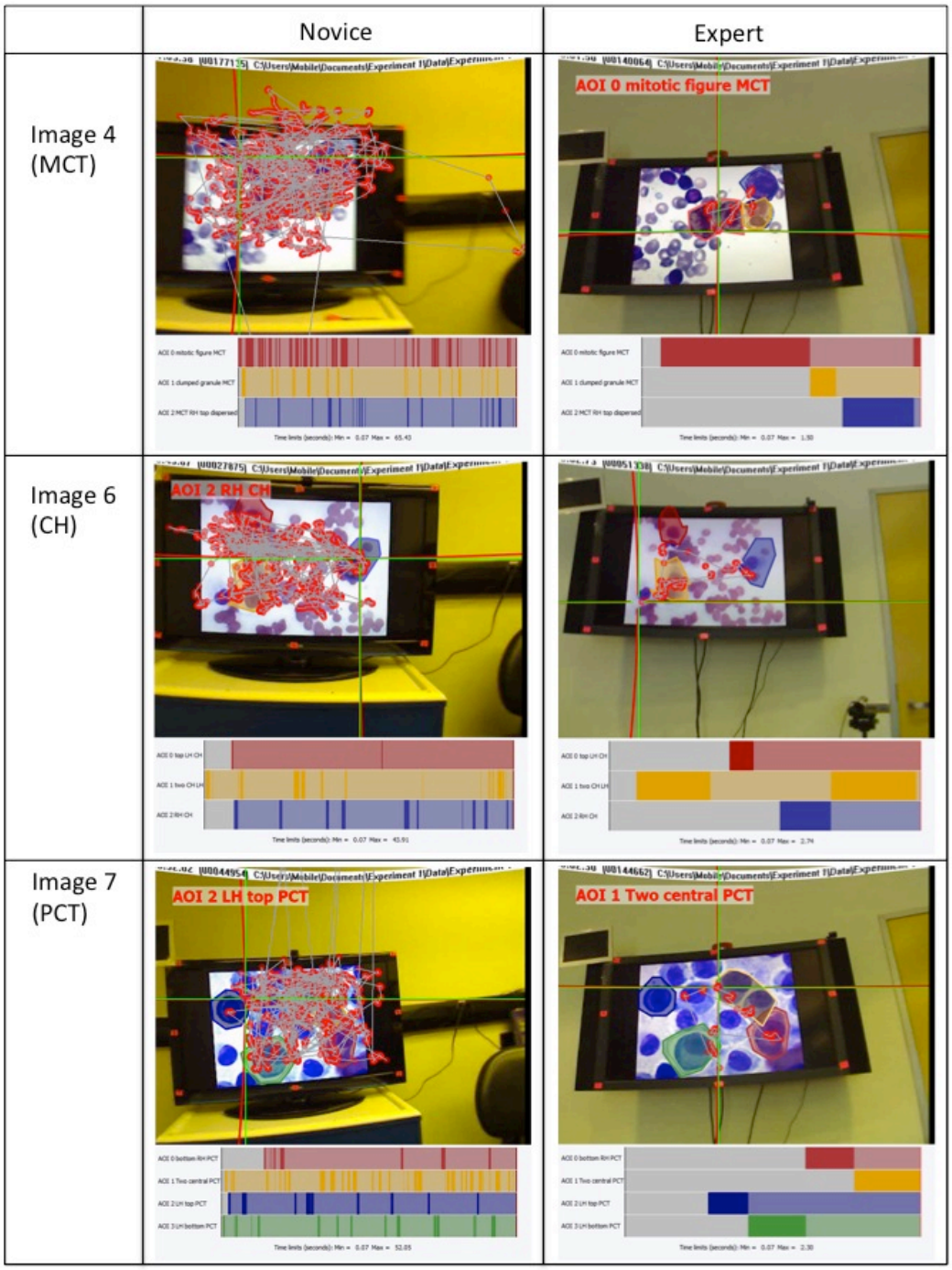


Figure 4.2 Eye-movement patterns for novices and experts for images 4 (mast cell tumor) (novice 13, expert 4), 6 (cutaneous histiocytoma) (novice 17, expert 8) and 7 (extramedullary plasmacytoma) (novice 23, expert 12).

4.3 Visual Reasoning Education Interventions

4.3.1 Descriptive statistics

Twenty-eight final year (fourth year) DVM students participated in the teaching intervention study of the research project, of which there were 20 (71%) female and 8 (29%) male students. Group 1: traditional teaching (n = 8) had 7 female and 1 male, Group 2: basic visual teaching (n = 10) had 6 female and 4 male, and Group 3: extended visual teaching intervention (n = 10) had 7 female and 3 male students. For pre-test image 1 for student 3 (Group 1), post-test images 6 to 10 for student 13 (Group 3), and post-test images 8 to 10 for student 2 (Group 1) there was eye-tracking recording failures. As such, pre-test image 1 and post-test image 6 (CH) were excluded from analysis for student 3, pre-test images 1 to 5 and post-test images 6 to 10 were excluded for student 13, and pre-test image 2 and post test image 10 (LSA), pre-test image 3 and post-test image 8 (TVT) and pre-test image 4 and post-test image 9 (MCT) were excluded for student 2 for eye-tracking data only. Diagnostic accuracy for all students was included.

Overall means and standard deviations, minimum and maximum scores, skewness and kurtosis for the each of the five corresponding pre- and post-test image diagnoses are tabulated: 1) cutaneous histiocytoma (Table 4.16), 2) cutaneous lymphoma (Table 4.17), 3) transmissible venereal tumor (Table 4.18), 4) mast cell tumor (Table 4.19) and 5) extramedullary plasmacytoma (Table 4.20), as well as for aggregated scores for the pre-test (images 1 to 5) and post-test (images 6 to 10) images (Table 4.21). Although multiple measures had a significantly non-normal distributions based on a K-S test ($p < 0.05$) parametric analyses were still judged appropriate. A summary of the means and *SDs* for time to diagnosis (s) and percentage time spent in the AOIs (%) for each slide is provided in Table 4.22. Diagnostic accuracy (i.e., correct

= 1, incorrect = 0), as an independent dichotomous variable, was recorded for each slide and total scores added to calculate % correct for individual participants and for each slide image (Table 4.22).

Table 4.16 Descriptive data for cutaneous histiocytoma (CH) pre- and post-test time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 26).

	Variable	**Min.**	**Max.**	**Mean (*SD*)**	**Skewness (*SE*)**	**Kurtosis (*SE*)**
Image 1 Pre-test CH	Time (s)*	10.4	101.7	50.5 (29.1)	0.7 (0.4)	-0.9 (0.9)
	% time in AOI	0.6	78.0	50.8 (19.9)	-0.7 (0.5)	0.3 (0.9)
Image 6 Post-test CH	Time (s)*	3.9	66.2	19.1 (13.4)	2.0 (0.5)	5.1 (0.9)
	% time in AOI*	0.0	87.0	65.1 (20.0)	-1.5 (0.5)	3.0 (0.9)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.17 Descriptive data for cutaneous lymphoma (LSA) pre- and post-test time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 26).

	Variable	**Min.**	**Max.**	**Mean (*SD*)**	**Skewness (*SE*)**	**Kurtosis (*SE*)**
Image 2 Pre-test LSA	Time (s)*	4.9	85.3	32.5 (21.6)	0.9 (0.5)	-2.8 (0.9)
	% time in AOI	28.0	81.6	53.5 (12.8)	-0.2 (0.5)	0.0 (0.9)
Image 10 Post-test LSA	Time (s)*	1.13	49.41	11.6 (11.5)	2.1 (0.5)	4.7 (0.9)
	% time in AOI	1.6	54.3	33.6 (13.1)	-0.9 (0.5)	1.2 (0.9)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.18 Descriptive data for transmissible venereal tumor (TVT) pre- and post-test time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 26).

	Variable	**Min.**	**Max.**	**Mean (*SD*)**	**Skewness (*SE*)**	**Kurtosis (*SE*)**
Image 3 Pre-test TVT	Time (s)	14.3	92.5	44.5 (30.8)	0.8 (0.5)	0.0 (0.9)
	% time in AOI	14.3	56.0	37.6 (9.0)	-0.5 (0.5)	0.9 (0.9)
Image 8 Post-test TVT	Time (s)*	1.9	13.3	5.6 (3.5)	1.1 (0.5)	0.1 (0.9)
	% time in AOI	0.0	93.5	58.5 (22.8)	-0.8 (0.5)	0.6 (0.9)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.19 Descriptive data for mast cell tumor (MCT) pre- and post-test time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 26).

	Variable	Min.	Max.	Mean (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)
Image 4 Pre-test MCT	Time (s)*	2.4	238.6	25.1 (45.0)	4.6 (0.5)	22.3 (0.9)
	% time in AOI*	12.4	92.0	59.1 (22.2)	-0.9 (0.5)	0.3 (0.9)
Image 9 Post-test MCT	Time (s)*	1.4	10.8	3.9 (2.4)	1.7 (0.5)	2.6 (0.9)
	% time in AOI	0.0	87.1	49.0 (21.6)	-0.1 (0.5)	0.0 (0.9)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.20 Descriptive data for extramedullary plasmacytoma (PCT) pre- and post-test time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 27).

	Variable	Min.	Max.	Mean (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)
Image 5 Pre-test PCT	Time (s)*	7.8	84.6	32.9 (20.6)	0.9 (0.5)	-0.1 (0.9)
	% time in AOI	1.8	67.4	37.4 (14.1)	-0.5 (0.5)	0.9 (0.9)
Image 7 Post-test PCT	Time (s)*	2.8	90.2	17.3 (19.5)	2.6 (0.5)	7.1 (0.9)
	% time in AOI	13.6	75.0	47.8 (13.5)	-0.2 (0.5)	1.4 (0.9)

* Using a K-S test for normalcy, these were significantly non-normal $p < 0.001$

Table 4.21 Descriptive data for the aggregated pre- (images 1-5) and post-test (images 6-10) time to diagnosis (s) and percentage time in AOI (%) for all teaching groups (n = 131).

	Variable	Min.	Max.	Mean (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)
Pre-test Images 1-5	Time (s)	2.4	238.6	36.7 (29.8)	2.8 (0.2)	15.2 (0.4)
	% time in AOI	0.6	92.0	47.5 (18.4)	0.0 (0.2)	-0.1 (0.4)
Post-test Images 6-10	Time (s)	1.13	90.22	11.6 (13.4)	3.0 (0.2)	11.5 (0.4)
	% time in AOI	0.0	93.46	50.8 (21.3)	-0.2 (0.2)	-0.3 (0.4)

Table 4.22 Mean and *SD* for the independent variables time to diagnosis (s), percentage time in AOI (%) and the diagnostic accuracy (percentage correct) for pre-test images 1-5 and post-test images 6-10.

	Image ¶	*n*	Time (s) *M* (*SD*)	% time in AOI *M* (*SD*)	Diagnostic accuracy (% correct)§
Pre-test	1 CH	26	48.0 (29.6)	50.6 (20.3)	0.0
	2 LSA	26	32.5 (21.6)	53.5 (12.8)	42.9
	3 TVT	26	44.5 (20.8)	37.6 (9.0)	0.0
	4 MCT	26	25.1 (45.0)	59.1 (22.2)	82.1
	5 PCT	27	32.9 (20.6)	37.4 (14.1)	17.9
Post-test	6 CH	26	19.3 (13.6)	65.5 (20.3)	67.9
	7 PCT	27	17.3 (19.5)	47.8 (13.5)	89.3
	8 TVT	26	5.6 (3.5)	58.5 (22.8)	96.4
	9 MCT	26	3.9 (2.4)	49.0 (21.6)	100.0
	10 LSA	26	11.6 (11.5)	33.6 (13.1)	96.4

¶CH= cutaneous histiocytoma; LSA = cutaneous lymphoma; TVT = transmissible venereal tumor; MCT = mast cell tumor; PCT = extramedullary plasmacytoma, § $n = 28$

4.3.2 Difference in teaching intervention over time

4.3.2.1 Time to diagnosis and percentage time in the AOIs

As this was a preliminary study we were interested in establishing the performance of each paired diagnosis image. Thus, for each diagnosis (cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma), for each teaching intervention group (1-3), and all teaching groups combined, paired *t*-tests were run to ascertain if there was a statistically significant difference between the pre- and post-test for time (s) to diagnosis and percentage time in AOI (%).

4.3.2.1.1 Cutaneous histiocytoma

For traditional teaching intervention Group one, there were no significant differences between the pre- and post-test for time to diagnosis and percentage time spent in the AOIs ($p > 0.017$).

For basic visual teaching intervention Group two, there were no significant differences between the pre- and post-test for time to diagnosis and percentage time spent in the AOIs ($p > 0.017$).

For extended visual teaching intervention Group three, there was a significant decrease between pre- and post-test for time to diagnosis ($M = 49.7$s, $SD = 31.0$s; $M = 17.5$s, $SD = 14.0$s, respectively), $t(8) = 3.234$, $p < 0.017$; but no significant difference between the pre- and post-test for percentage time in the AOI ($p > 0.017$).

When all three teaching intervention groups (1-3) were combined, there was a significant decrease between the pre- and post test for time to diagnosis ($M = 48.5$s, $SD = 29.6$s; $M = 19.3$s, $SD = 13.6$s, respectively), $t(26) = 4.816$, $p < 0.017$ and a significant increase in the percentage time spent in the AOIs ($M = 50.6\%$, $SD = 20.3\%$; $M = 65.5\%$, $SD = 20.3\%$, respectively), $t(26) = -2.742$, $p < 0.017$ (Table 4.23).

Table 4.23 Time to diagnosis (s) and percentage time spent in the AOIs (%) for the paired pre- and post-test for cutaneous histiocytoma for teaching intervention Groups 1, 2, and 3 and all groups combined.

Teaching intervention group	Parameter	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Mean difference (*SD*)	*t*-value	*df*	*p*-	Effect size Cohen's *d* ‡ (95% CI)
1 (*n* = 7)	Time (s)	48.0 (20.9)	28.5 (17.9)	19.5 (21.9)	2.365	6	0.056	1.79 (0.04 - 3.54)
	% time in AOI	46.2 (28.1)	65.7 (19.5)	- 19.48 (40.4)	-1.276	6	0.249	-0.96 (-2.53 - 0.60)
2 (*n* = 10)	Time (s)	47.7 (35.9)	14.6 (6.1)	33.0 (37.8)	2.762	9	0.022	1.75 (0.29 – 3.20)
	% time in AOI	52.4 (20.3)	64.4 (14.8)	-12.1 (13.8)	-2.777	9	0.022	-1.76 (-3.22 - -0.30)
3 (*n* = 9)	Time (s)	49.7 (31.0)	17.5 (14.0)	32.2 (29.9)	3.234	8	0.012*	2.16 (0.51 – 3.80)
	% time in AOI	51.9 (14.3)	66.5 (27.4)	-14.6 (30.8)	-1.422	8	0.193	-0.95 (-2.33 – 0.43)
All groups (*n* = 26)	Time (s)	48.5 (29.6)	19.3 (13.6)	29.1 (30.8)	4.816	25	<0.001†	1.89 (0.96 – 2.81
	% time in AOI	50.6 (20.3)	65.5 (20.3)	-14.9 (27.8)	-2.742	25	0.011*	-1.08 (-1.90 - -0.25)

* $p < 0.017$, † $p < 0.001$. ‡ Cohen's $d > 0.80$ large effect size difference.

4.3.2.1.2 Cutaneous lymphoma

For traditional teaching intervention Group one, there was no significant difference between the pre- and post-test for time to diagnosis and percentage time spent in the AOIs ($p > 0.017$).

For basic visual teaching intervention Group two, there was no significant difference between the pre- and post-test for time to diagnosis ($p > 0.017$) and a significant decrease between the pre- and post-test in the percentage time spent in the AOI ($M = 52.2\%$, $SD = 16.0\%$; $M = 29.9\%$, $SD = 13.9\%$, respectively), $t(9) = 4.404$, $p < 0.017$. .

For extended visual teaching intervention Group three, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 41.7$s, $SD = 22.6$s; $M = 7.4$s, $SD = 5.4$s,

respectively), $t(8) = 4.670$, $p < 0.017$ and a significant decrease in the percentage time spent in the AOIs ($M = 54.6\%$, $SD = 11.6\%$; $M = 36.0\%$, $SD = 15.4\%$, respectively), $t(8) = 5.534$, $p < 0.017$

When all three teaching intervention groups (1 to 3) were combined, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 32.5$s, $SD = 21.6$s; $M = 11.6$s, $SD = 11.5$s, respectively), $t(25) = 4.591$, $p < 0.017$ and a significant decrease in the percentage time in the AOIs ($M = 53.5\%$, $SD = 12.8\%$; $M = 33.6\%$, $SD = 13.1\%$, respectively, $t(25) = 7.296$, $p < 0.017$.

Table 4.24 Time to diagnosis (s) and percentage time spent in the AOIs (%) for the paired pre- and post-test for cutaneous lymphoma for teaching intervention Groups 1, 2, and 3 and all groups combined.

Teaching intervention group	**Parameter**	**Pre-test *M* (*SD*)**	**Post-test *M* (*SD*)**	**Mean difference (*SD*)**	***t*-value**	***df***	***P***	**Effect size Cohen's *d* ‡ (95% CI)**
1 ($n = 7$)	Time (s)	27.1 (14.9)	13.8 (16.7)	13.2 (25.5)	1.375	6	0.218	1.04 (-0.54 – 2.62)
	% time in AOI	53.8 (10.9)	35.9 (8.5)	17.9 (16.2)	2.924	6	0.026	2.21 (0.33 – 4.09)
2 ($n = 10$)	Time (s)	28.1 (23.6)	13.8 (11.4)	14.2 (18.9)	2.374	9	0.042	1.50 (0.10 – 2.90)
	% time in AOI	52.2 (16.0)	29.9 (13.9)	22.3 (16.0)	4.404	9	0.002*	2.79 (1.05 – 4.53)
3 ($n = 9$)	Time (s)	41.7 (22.6)	7.4 (5.4)	34.3 (22.1)	4.670	8	0.002*	3.11 (1.17 – 5.06)
	% time in AOI	54.6 (11.6)	36.0 (15.4)	18.5 (10.0)	5.534	8	0.001†	3.69 (1.54 – 5.84)
All groups ($n = 26$)	Time (s)	32.5 (21.6)	11.6 (11.5)	20.9 (23.2)	4.591	25	<0.001†	1.80 (0.90 – 2.71)
	% time in AOI	53.5 (12.8)	33.6 (13.1)	19.8 (13.9)	7.296	25	<0.001†	2.86 (1.77 – 4.00)

*$p < 0.017$, †$p < 0.001$. ‡ Cohen's $d > 0.80$ large effect size difference.

4.3.2.1.3 Transmissible venereal tumor

For traditional teaching intervention Group one, there was no significant difference between the pre- and post-test for the percentage time spent in the AOI ($p > 0.017$), but a significant decrease between the pre- and post-test for time to diagnosis ($M = 42.1$s, $SD = 15.2$s; $M = 8.3$s, $SD = 3.0$s, respectively), $t(6) = 6.020$, $p < 0.017$.

For basic visual teaching intervention Group two, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 50.0$s, $SD = 27.4$s; $M = 5.3$s, $SD = 3.2$s, respectively), $t(9) = 5.279$, $p < 0.017$; a significant increase in the percentage time spent in the AOIs ($M = 37.4\%$, $SD = 11.6\%$; $M = 58.5\%$, $SD = 14.7\%$, respectively), $t(9) = -4.680$, $p < 0.017$.

For extended visual teaching intervention Group three, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 40.2$s, $SD = 16.4$s; $M = 3.8$s, $SD = 3.0$s, respectively), $t(8) = 6.799$, $p < 0.017$. No significant difference was noted for the percentage time spent in the AOIs, $p > 0.017$.

When all three teaching intervention groups (1-3) were combined, there was a significant decrease between the pre- and post test for time to diagnosis ($M = 44.5$s, $SD = 20.8$s; $M = 5.6$s, $SD = 3.5$s, respectively), $t(25) = 9.721$, $p < 0.017$, and a significant increase in the percentage time spent in the AOIs ($M = 37.6\%$, $SD = 9.0\%$; $M = 58.5\%$, $SD = 22.8\%$, respectively), $t(25) = -4.190$, $p < 0.017$ (Table 4.25).

Table 4.25 Time to diagnosis (s) and percentage time spent in the AOIs (%) for the paired pre- and post-test for transmissible venereal tumor for teaching intervention Groups 1, 2, and 3 and all groups combined.

Teaching intervention group	Parameter	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Mean difference (*SD*)	*t*-value	*df*	*p*	Effect size Cohen's *d* ‡ (95% CI)
1 (*n* = 7)	Time (s)	42.1 (15.2)	8.3 (3.0)	33.8 (14.9)	6.020	6	0.001†	4.55 (1.74 - 7.36)
	% time in AOI	36.7 (8.0)	53.3 (27.6)	-16.6 (34.0)	-1.289	6	0.245	-0.97 (-2.54 - 0.59)
2 (*n* = 10)	Time (s)	50.0 (27.4)	5.3 (3.2)	44.7 (26.7)	5.279	9	0.001†	3.34 (1.42 – 5.26)
	% time in AOI	37.4 (11.6)	58.5 (14.7)	-21.1 (14.2)	-4.680	9	0.001†	-2.96 (-4.75 - -1.17)
3 (*n* = 9)	Time (s)	40.2 (16.4)	3.8 (3.0)	36.5 (16.1)	6.799	8	<0.001†	4.53 (2.06 – 7.00)
	% time in AOI	38.4 (7.1)	62.5 (27.6)	-24.1 (30.0)	-2.411	8	0.042	-1.61 (-3.11 - -0.10)
All groups (*n* = 26)	Time (s)	44.5 (20.8)	5.6 (3.5)	38.9 (20.4)	9.721	25	<0.001†	3.81 (2.52 – 5.10)
	% time in AOI	37.6 (9.0)	58.5 (22.8)	-20.9 (25.5)	-4.190	25	<0.001†	-1.64 (-2.53 - -0.75)

*$p < 0.017$,†$p < 0.001$.‡ Cohen's $d > 0.80$ large effect size difference.

4.3.2.1.4 Mast cell tumor

For traditional teaching intervention Group one, there was no significant difference between the pre- and post-test for time to diagnosis and for the percentage time spent in the AOI ($p > 0.017$).

For basic visual teaching intervention Group two, there was no significant difference between the pre- and post-test for the time to diagnosis and percentage time spent in the AOI ($p > 0.017$).

For extended visual teaching intervention Group three, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 17.3$s, $SD = 10.7$s; $M = 2.5$s, $SD = 0.8$s,

respectively), $t(8) = 3.983$, $p < 0.017$.but no significant difference between the pre- and post-test for the percentage time spent in the AOI ($p > 0.017$).

When all three teaching intervention groups (1-3) were combined, there was no significant difference between the pre- and post-test for time to diagnosis and for the percentage time spent in the AOI ($p > 0.017$) (Table 4.26).

Table 4.26 Time to diagnosis (s) and percentage time spent in the AOIs (%) for the paired pre- and post-test for mast cell tumor for teaching intervention Groups 1, 2, and 3 and all groups combined.

Teaching intervention group	Parameter	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Mean difference (*SD*)	*t*-value	*df*	*p*	Effect size Cohen's *d* ‡ (95% CI)
1 ($n = 7$)	Time (s)	15.3 (10.7)	5.9 (3.3)	9.4 (10.0)	2.484	6	0.048	1.88 (0.10 – 3.66)
	% time in AOI	60.8 (22.2)	45.5 (21.5)	15.3 (31.9)	1.266	6	0.252	0.96 (-0.61 – 2.52)
2 ($n = 10$)	Time (s)	39.1 (71.4)	3.7 (1.5)	35.4 (72.0)	1.556	9	0.154	0.98 (-0.33 – 2.30)
	% time in AOI	55.7 (29.0)	49.4 (16.5)	6.3 (34.2)	0.579	9	0.577	0.37 (-0.88 – 1.62)
3 ($n = 9$)	Time (s)	17.3 (10.7)	2.5 (0.8)	14.8 (11.1)	3.983	8	0.004*	2.66 (0.86 – 4.45)
	% time in AOI	61.6 (14.0)	51.3 (28.1)	10.3 (34.2)	0.901	8	0.394	0.60 (-0.74 – 1.94)
All groups ($n = 26$)	Time (s)	25.1 (45.0)	3.9 (2.4)	21.3 (45.5)	2.386	25	0.025	0.94 (0.13 – 1.75)
	% time in AOI	59.1 (22.2)	49.0 (21.6)	10.1 (32.4)	1.584	25	0.126	0.62 (-0.17 – 1.41)

*$p < 0.017$.‡ Cohen's $d > 0.80$ large effect size difference.

.

4.3.2.1.5 Extramedullary plasmacytoma

For traditional teaching intervention Group one, there was no significant difference between the pre- and post-test for time to diagnosis ($p > 0.017$) and percentage time in AOI ($p > 0.017$).

For basic visual teaching intervention Group two, there was no significant difference between the pre- and post test for time to diagnosis ($p > 0.017$), but a significant increase between the pre- and post-test for percentage time spent in the AOIs ($M = 29.6\%$, $SD = 15.5\%$; $M = 48.0\%$, $SD = 14.7\%$, respectively), $t(9) = -4.517$, $p < 0.017$

For extended visual teaching intervention Group three, there was a significant decrease between the pre- and post test for time to diagnosis ($M = 35.8$s, $SD = 18.4$s; $M = 11.3$s, $SD = 11.3$s, respectively), $t(8) = 4.817$, $p < 0.017$, but no significant difference between the pre- and post test for percentage time spent in the AOIs ($p > 0.017$).

When all three teaching intervention groups (1-3) were combined, there was a significant decrease between the pre- and post-test for time to diagnosis ($M = 32.9$s, $SD = 20.6$s; $M = 17.3$s, $SD = 19.5$s, respectively), $t(26) = 2.627$, $p < 0.017$ and a significant increase in the percentage time spent in the AOI ($M = 37.4\%$, $SD = 14.1\%$; $M = 47.8\%$, $SD = 13.5\%$, respectively), $t(26) = -3.586$, $p < 0.017$ (Table 4.27).

Table 4.27 Time to diagnosis (s) and percentage time spent in the AOIs (%) for the paired pre- and post-test for extramedullary plasmacytoma for teaching intervention Groups 1, 2, and 3 and all groups combined.

Teaching intervention group	Parameter	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Mean difference (*SD*)	*t*-value	*df*	*p*	Effect size Cohen's *d* ‡ (95% CI)
1 (*n* = 8)	Time (s)	29.5 (22.1)	17.6 (15.1)	11.9 (30.3)	1.112	7	0.303	0.79 (-0.65 – 2.23)
	% time in AOI	43.3 (14.5)	45.2 (4.5)	-1.92 (14.2)	-0.385	7	0.712	-0.27 (-1.66 – 1.12)
2 (*n* = 10)	Time (s)	33.0 (23.1)	22.4 (27.3)	10.6 (41.4)	0.807	9	0.440	0.51 (-0.75 – 1.77)
	% time in AOI	29.6 (15.5)	48.0 (14.7)	-18.4 (12.9)	-4.517	9	0.001†	-2.86 (-4.62 – -1.09)
3 (*n* = 9)	Time (s)	35.8 (18.4)	11.3 (11.3)	24.5 (15.3)	4.817	8	0.001†	3.21 (1.23 – 5.19)
	% time in AOI	40.7 (8.4)	49.9 (17.8)	-9.2 (15.1)	-1.821	8	0.106	-1.21 (-2.64 – 0.21)
All groups (*n* = 27)	Time (s)	32.9 (20.6)	17.3 (19.5)	15.6 (30.9)	2.627	26	0.014*	1.01 (0.21 – 1.81)
	% time in AOI	37.4 (14.1)	47.8 (13.5)	-10.5 (15.2)	-3.586	26	0.001†	-1.38 (-2.22 - -0.54)

*$p < 0.017$,†$p < 0.001$. ‡ Cohen's $d > 0.80$ large effect size difference.

4.3.2.2 Diagnostic accuracy

For each diagnosis (cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma), for each teaching intervention Group (1-3), and all teaching groups combined, chi-square tests were performed to ascertain if there was a difference between the pre- an post-test for diagnostic accuracy (percentage correct) (Table 4.28).

For cutaneous histiocytoma there was a significant increase in the percentage correct for teaching intervention Group 1 (χ^2 (1, N = 8) = 7.27, $p < 0.017$), teaching intervention Group 2 (χ^2

(1, $N = 10$) = 8.57, $p < 0.017$), teaching intervention Group 3 (χ^2 (1, $N = 10$) = 13.33, $p < 0.017$) and all groups combined (χ^2 (1, $N = 28$) = 28.76, $p < 0.017$).

For cutaneous lymphoma there was a significant increase in the percentage correct for teaching intervention Group 2 (χ^2 (1, $N = 10$) = 8.57, $p < 0.017$), teaching intervention Group 3 (χ^2 (1, $N = 10$) = 8.57, $p < 0.017$) and all groups combined (χ^2 (1, $N = 28$) = 19.01, $p < 0.017$) but no significant difference between the pre and post-test for teaching intervention Group 1 ($p > 0.017$).

For transmissible venereal tumor there was a significant increase in the percentage correct for teaching intervention Group 1 (χ^2 (1, $N = 8$) = 12.44, $p < 0.017$), teaching intervention Group 2 (χ^2 (1, $N = 10$) = 20.00, $p < 0.017$), teaching intervention Group 3 (χ^2 (1, $N = 10$) = 20.00, $p < 0.017$) and all groups combined (χ^2 (1, $N = 28$) = 52.14, $p < 0.017$).

For mast cell tumor there were no significant differences between the pre-test and post-test percentage correct for teaching intervention Group 1 ($p > 0.017$), teaching intervention Group 2 ($p > 0.017$) and teaching intervention Group 3 ($p > 0.017$) but a significant increase in percentage correct when all groups were combined (χ^2 (1, $N = 28$) = 5.49, $p < 0.017$).

For extramedullary plasmacytoma there was a significant increase in the percentage correct for teaching intervention Group 2 (χ^2 (1, $N = 10$) = 12.8, $p < 0.017$), teaching intervention Group 3 (χ^2 (1, $N = 10$) = 16.36, $p < 0.017$) and all groups combined (χ^2 (1, $N = 28$) = 28.72, $p < 0.017$) but no significant difference between pre- and post-test scores for teaching intervention Group 1 ($p > 0.017$).

Table 4.28 Chi-square test comparing the diagnostic accuracy (percent correct) for the paired pre- and post-test diagnoses, for teaching intervention Groups 1, 2, and 3 and all groups combined.

Diagnosis	Teaching intervention group	*n*	Pre-test % correct	Post-test % correct	Percent diff. (%)	*p*
Cutaneous histiocytoma (CH)	1	8	0	62.5	62.5	0.007*
	2	10	0	60.0	60.0	0.003*
	3	10	0	80.0	80.0	<0.001†
	All groups	28	0	67.8	67.8	<0.001†
Cutaneous lymphoma (LSA)	1	8	50.0	87.5	37.5	0.106
	2	10	40.0	100.0	60.0	0.003*
	3	10	40.0	100.0	60.0	0.003*
	All groups	28	42.8	96.4	53.6	<0.001†
Transmissible venereal tumor (TVT)	1	8	0	87.5	87.5	<0.001†
	2	10	0	100.0	100.0	<0.001†
	3	10	0	100.0	100.0	<0.001†
	All groups	28	0	96.4	96.4	<0.001†
Mast cell tumor (MCT)	1	8	87.5	100.0	12.5	0.302
	2	10	80	100.0	20.0	0.136
	3	10	80	100.0	20.0	0.136
	All groups	28	82.1	100.0	17.9	0.019
Extramedullary plasmacytoma (PCT)	1	8	37.5	75.0	37.5	0.131
	2	10	10.0	90.0	80.0	<0.001†
	3	10	10.0	90.0	80.0	<0.001†
	All groups	28	17.9	89.3	71.4	<0.001†

*$p < 0.017$, †$p < 0.001$.

4.3.3 Differences between teaching groups for the individual diagnoses

A one-way between-group analysis of covariance (ANCOVA) was used to compare the effectiveness of teaching interventions 1, 2 and 3 for improving visual diagnostic reasoning skills on each of the paired scores for the 5 pre and 5 post-test images (cutaneous histiocytoma [CH], cutaneous lymphoma [LSA], transmissible venereal tumor [TVT], mast cell tumor [MCT] and plasmacytoma [PCT]). The independent variable was the teaching intervention groups (1-3).

The dependent variables tested were post-test time to diagnosis and post-test percentage time spent in the AOIs. The pre-test scores for time to diagnosis and the percentage time in the AOIs were used as covariates in the analysis. Preliminary tests for the violation of assumptions were conducted, including tests for normality, linearity, homogeneity of variances and homogeneity of regression slopes. A post-test Tukey analysis was performed when significant differences were observed.

Using the scores for diagnostic accuracy (correct = 1, incorrect = 0) for the post-test for each diagnosis (cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor and extramedullary plasmacytoma), the three teaching groups (1-3) were compared using a chi-square test (Table 4.34). A chi-square test was not performed for the mast cell tumor as the post-test percentage correct was constant for all three teaching intervention groups.

For the cutaneous histiocytoma, after adjusting for pre-test scores, no significant difference was found between the three teaching groups for time to diagnosis ($p > 0.017$) and percentage time spent in AOIs ($p > 0.017$) (Table 4.29). No significant difference was noted between the three teaching intervention groups for diagnostic accuracy (percent correct) ($p > 0.017$) (Table 4.34).

For the cutaneous lymphoma, after adjusting for pre-test scores, no significant difference was found between the three teaching groups for time to diagnosis ($p > 0.017$) and percentage time spent in AOIs ($p > 0.017$) (Table 4.30). No significant difference was found between the three teaching intervention groups for diagnostic accuracy (percent correct) ($p > 0.017$) (Table 4.34)

For the transmissible venereal tumor, after adjusting for pre-test scores, no significant difference was found between the three teaching groups for time to diagnosis ($p > 0.017$) and

percentage time spent in AOIs ($p > 0.017$) (Table 4.31). No significant difference between the three teaching intervention groups for diagnostic accuracy (percent correct) was noted ($p > 0.017$) (Table 4.34).

For the mast cell tumor, after adjusting for pre-test scores, a significant difference was found between the three teaching groups for time to diagnosis F (2, 22) = 5.965, $p < 0.017$, partial eta squared = 0.352. A Tukey's post-test indicated that the mean time to diagnosis for teaching intervention Group 1 (M = 5.9s, SD = 0.8s) was significantly longer than the mean time to diagnosis for teaching intervention Group 3 (M = 2.5s, SD = 0.7s). No significant difference was found between the three teaching groups for percentage time spent in the AOIs ($p > 0.017$). Diagnostic accuracy between the three teaching intervention group's post-test results for the mast cell tumor was constant.

For the extramedullary plasmacytoma, after adjusting for pre-test scores, no significant difference was found between the three teaching groups for time to diagnosis ($p > 0.017$) and the percentage time spent in the AOIs ($p > 0.017$) (Table 4.33). No significant difference between the three teaching intervention groups for diagnostic accuracy (percent correct) was noted ($p > 0.017$) (Table 4.34).

Table 4.29 ANCOVA results comparing differences in Groups 1-3 for cutaneous histiocytoma results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	*p*	Partial Eta squared
Time (s)	1	7	28.5 (4.9)	2	2.467	0.108	0.183
	2	10	14.7 (4.1)				
	3	9	17.4 (4.3)				
% time in AOI	1	7	66.0 (8.2)	2	0.025	0.975	0.002
	2	10	64.3 (6.8)				
	3	9	66.4 (7.2)				

Table 4.30 ANCOVA results comparing differences in Groups 1-3 for cutaneous lymphoma results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	*p*	Partial Eta squared
Time (s)	1	7	14.5 (4.4)	2	1.334	0.284	0.108
	2	10	14.4 (3.7)				
	3	9	6.3 (4.0)				
% time in AOI	1	7	35.8 (4.6)	2	0.556	0.581	0.048
	2	10	30.4 (3.9)				
	3	9	35.6 (4.1)				

Table 4.31 ANCOVA results comparing differences in Groups 1-3 for transmissible venereal tumor results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	Sig. *p*	Partial Eta squared
Time (s)	1	7	8.4 (1.2)	2	4.412	0.024	0.286
	2	10	5.3 (1.0)				
	3	9	3.9 (1.0)				
% time in AOI	1	7	53.0 (9.0)	2	0.336	0.718	0.030
	2	10	58.4 (7.5)				
	3	9	62.8 (7.9)				

Table 4.32 ANCOVA results comparing differences in Groups 1-3 for mast cell tumor results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	Sig. *p*	Partial Eta squared
Time (s)	1	7	5.9 (0.8)	2	5.965	0.009*	0.352
	2	10	3.8 (0.6)				
	3	9	2.5 (0.7)				
% time in AOI	1	7	45.7 (8.6)	2	0.130	0.879	0.012
	2	10	49.1 (7.3)				
	3	9	51.6 (7.6)				

*$p < 0.017$

Table 4.33 ANCOVA results comparing differences in Groups 1-3 for extramedullary plasmacytoma results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	Sig. *p*	Partial Eta squared
Time (s)	1	8	17.1 (7.0)	2	0.691	0.511	0.057
	2	10	22.5 (6.3)				
	3	9	11.7 (6.6)				
% time in AOI	1	8	42.3 (4.6)	2	1.072	0.359	0.085
	2	10	51.8 (4.3)				
	3	9	48.3 (4.2)				

Table 4.34 Chi-squared results comparing Groups 1-3 for post-test percentage correct (%) for each of the diagnoses cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma.

Diagnosis	Teaching intervention Group	*n*	Post-test % correct	*df*	χ^2	*p*
Cutaneous histiocytoma	1	8	62.5	2	1.064	0.587
	2	10	60.0			
	3	10	80.0			
Cutaneous lymphoma	1	8	87.5	2	2.593	0.274
	2	10	100.0			
	3	10	100.0			
Transmissible venereal tumor	1	8	87.5	2	2.593	0.274
	2	10	100.0			
	3	10	100.0			
Mast cell tumor§	1	8	100.0	-	-	-
	2	10	100.0			
	3	10	100.0			
Extramedullary plasmacytoma	1	8	75.0	2	2.912	0.233
	2	10	90.0			
	3	10	90.0			

§ Chi-square test was not performed for this group as the post-test percentage correct is constant.

4.3.4 Differences between teaching groups for aggregated pre- and post-test results

A one-way between-group analysis of covariance (ANCOVA) was run to compare the effectiveness of teaching interventions 1, 2 and 3 for improving visual diagnostic reasoning skills on aggregated scores for the five pre and five post-test images (Table 4.35). Using the aggregated scores for diagnostic accuracy (correct = 1, incorrect = 0) for the post-tests combined (cutaneous histiocytoma, cutaneous lymphoma, mast cell tumor, transmissible venereal tumor and extramedullary plasmacytoma), the three teaching groups (1-3) were compared using a chi-square test (Table 4.36).

After adjusting for pre-test scores, no significant difference was found between the three teaching groups for time to diagnosis ($p > 0.017$) and percentage time spent in the AOIs ($p >$

0.017) (Table 4.35). There was no significant difference between teaching intervention groups (1-3) for diagnostic accuracy (post-test percentage correct) ($p > 0.017$) (Table 4.36).

Table 4.35 ANCOVA results comparing differences in Groups 1-3 for aggregated results of time to diagnosis (s) and percentage time in AOI (%).

Variable	Teaching intervention group	*n*	Estimated Marginal Mean (*SD*)	*df*	*F*	*p*	Partial Eta squared
Time (s)	1	36	15.0 (2.2)	2	2.489	0.087	0.038
	2	50	11.9 (1.9)				
	3	45	8.5 (2.0				
% time in AOI	1	36	49.0 (3.6)	2	0.460	0.632	0.007
	2	50	50.0 (3.0)				
	3	45	53.3 (3.2)				

Table 4.36 Chi-square test comparing teaching intervention Groups 1-3 for post-test diagnostic accuracy (percent correct) for all diagnoses combined.

Diagnosis	Teaching intervention Group	*n*	Post-test % correct	*df*	χ^2	*p*
Aggregated post-test % correct	1	40	82.5	2	4.500	0.105
	2	50	90.0			
	3	50	96.0			

4.3.4.1 Qualitative differences in eye-movement patterns pre- and post-test between teaching intervention groups

In all three teaching groups, the pre-test eye-movement patterns were more complicated, with a higher number of fixations and saccades that formed a more complex eye-movement pattern compared to the post-test. In all cases, except for the MCT, pre-test eye-movement patterns were less focused on the AOIs and included multiple fixations in areas outside of the AOIs. For all groups at post-test, the patterns of eye-movement generally were more simplified

with fewer fixations and saccades and there was a pattern of movements that was more concentrated on the AOIs (Appendix 9a, 9b, and 9c). This was most apparent with students in Group 3 (Appendix 9c), and to a lesser degree students in Group 2 (Appendix 9b), with smoother eye-movement transitions and more attention focused on the AOIs. For students in Group 1, the difference between eye-movement patterns pre- and post-test for the five parallel diagnoses was subtle with little difference notable for several of the diagnoses (Appendix 9a).

4.3.5 Comparison of post-test performance for teaching intervention groups with expert performance (Objective 1)

4.3.5.1.1 Individual diagnoses

The time to diagnosis (s) and the percentage time spent in the AOIs (%) for the post-test for each diagnosis (cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma) for teaching intervention groups 1-3, were compared to the results for the experts (Objective 1) on each of the corresponding slides from objective 1 (images 6-10) using an independent-samples *t*-tests (Table 4.37). The diagnostic accuracy for the post-test for each diagnosis (cutaneous histiocytoma, cutaneous lymphoma, transmissible venereal tumor, mast cell tumor and extramedullary plasmacytoma) for teaching intervention groups 1-3, were compared to the results for the experts (Objective 1) on each of the corresponding slides (images 6-10) using chi-square test (Table 4.38).

For the post-test cutaneous histiocytoma (image 6), students in traditional teaching intervention Group one (M = 28.4s, SD = 17.9s) took significantly longer than the experts (M = 6.1s, SD = 2.1s) to reach a diagnosis $t(11) = 3.020$, $p < 0.017$ and students in basic visual teaching intervention Group two (M = 14.6s, SD = 6.1s) took significantly longer than experts

(M = 6.1s, SD = 2.1s) to reach a diagnosis $t(14) = 3.298$, $p < 0.017$. There was no significant difference between students in extended visual teaching intervention Group three and experts in the time to reach a diagnosis $p > 0.017$. There was no significant difference for percentage time spent in the AOIs between the experts and teaching intervention Group one ($p > 0.017$), the experts and Group two ($p > 0.017$), and the experts and teaching intervention Group three ($p > 0.017$) (Table 4.37). There was no significant difference between teaching intervention Groups one, two and three and the experts for diagnostic accuracy ($p > 0.017$) (Table 4.38).

For the post-test cutaneous lymphoma (image 10), there was no significant difference for time to diagnosis and percentage time spent in the AOIs between the experts and traditional teaching intervention Group one ($p > 0.017$), the experts and basic visual teaching intervention Group two ($p > 0.017$) and the experts and extended visual teaching intervention Group three ($p > 0.017$) (Table 4.37). There was no significant difference between teaching intervention Groups one, two and three and the experts for diagnostic accuracy ($p > 0.017$) (Table 4.38).

For the post-test transmissible venereal tumor (image 8), there was no significant difference for time to diagnosis and the percentage time spent in the AOIs between the experts and traditional teaching intervention Group one ($p > 0.017$), the experts and basic visual teaching intervention Group two ($p > 0.017$) and the experts and extended visual teaching intervention Group three ($p > 0.017$) (Table 4.37). There was no significant difference between teaching intervention groups one, two and three and the experts for diagnostic accuracy ($p > 0.017$) (Table 4.38).

For the post-test mast cell tumor (image 9), students in traditional teaching intervention Group one (M = 5.9s, SD = 3.3s) took significantly longer than the experts (M = 1.6s, SD = 1.0s) to reach a diagnosis $t(7.4) = 3.252$, $p < 0.017$; students in basic visual teaching intervention

Group two (M = 3.7s, SD = 1.5s) took significantly longer than experts (M = 1.6s, SD = 1.0s) to reach a diagnosis $t(14) = 2.880$, $p < 0.017$. There was no significant difference between students in the extended visual teaching intervention Group three and experts for the time to reach a diagnosis ($p > 0.017$). There was no significant difference for percentage time spent in the AOIs between the experts and traditional teaching intervention Group one ($p > 0.017$), the experts and basic visual teaching intervention Group two ($p > 0.017$), and the experts and extended visual teaching intervention Group three ($p > 0.017$). (Table 4.37). The post-test diagnostic accuracy scores for teaching intervention groups one, two and three were identical to the expert scores (Table 4.38).

For the post-test extramedullary plasmacytoma (image 7), there was no significant difference for time to diagnosis between the experts and traditional teaching intervention Group one ($p > 0.017$), the experts and basic visual teaching intervention Group two ($p > 0.017$), and the experts and extended visual teaching intervention Group three ($p > 0.017$). Students in traditional teaching intervention Group 1 (M = 45.2%, SD = 4.5%) spent on average less time in the AOIs than experts (M = 60.8%, SD = 9.8%) $t(6.6) = -3.617$, $p < 0.017$. There was no significant difference between students in teaching intervention Groups two and three and experts for the percentage time spent in the AOIs ($p > 0.017$). There was no significant difference between teaching intervention Groups 1 and 2 and the experts for diagnostic accuracy ($p > 0.017$) (Table 4.38). The post-test diagnostic accuracy scores for teaching intervention Group 3 were identical to the expert scores.

Table 4.37 Independent sample *t*-test comparing post-test teaching intervention Groups 1, 2 and 3 to the corresponding results from the expert group (n = 6) (Objective 1) for time to diagnosis (s) and percentage time spent in the AOIs (%) for each diagnosis (cutaneous histiocytoma [CH] (image 6), cutaneous lymphoma [LSA] (image 10), transmissible venereal tumor [TVT] (image 8), mast cell tumor [MCT] (image 9) and extramedullary plasmacytoma [PCT] (image 7).

Diagnosis (Image)	**Teaching intervention group**	**Parameter**	**Post-test *M* (*SD*)**	**Expert *M* (*SD***	***t*-value**	***df***	***p*-value**	**Effect size Cohen's *d* ‡ (95% CI)**
CH (Image 6)	1 (n = 7)	Time (s)	28.4 (17.9)	6.1 (2.1)	3.020	11	0.012*	1.68 (0.41 – 2.95)
		% time in AOI	65.7 (19.5)	74.5 (8.4)	-1.026	11	0.327	-0.57 (-1.68 – 0.54)
	2 (n = 10)	Time (s)	14.6 (6.1)	6.1 (2.1)	3.298	14	0.005*	1.70 (0.53 – 2.87)
		% time in AOI	64.4 (14.8)	74.5 (8.4)	-1.518	14	0.151	-0.78 (-1.83 – 0.26)
	3 (n = 9)	Time (s)§	17.5 (14.0)	6.1 (2.1)	2.403	8.5	0.041	1.27 (0.14 – 2.39)
		% time in AOI	66.5 (27.4)	74.5 (8.4)	-0.668	13	0.503	-0.35 (-1.39 – 0.69)
LSA (Image 10)	1 (n = 7)	Time (s)	13.8 (16.7)	7.0 (5.6)	0.952	11	0.361	0.53 (-0.58 – 1.64)
		% time in AOI	35.9 (8.5)	41.7 (16.1)	0.166	11	0.426	0.09 (-1.00 – 1.18)
	2 (n = 10)	Time (s)	13.8 (11.4)	7.0 (5.6)	1.350	14	0.198	0.70 (-0.34 – 1.74)
		% time in AOI	29.9 (13.8)	41.7 (16.1)	-1.549	14	0.144	-0.80 (-1.85 – 0.25)
	3 (n = 9)	Time (s)	7.4 (5.4)	7.0 (5.6)	0.110	13	0.914	0.06 (-0.98 – 1.10)
		% time in AOI	36.0 (15.4)	41.7 (16.1)	-0.681	13	0.508	-0.36 (-1.40 – 0.68)
TVT (Image 8)	1 (n = 7)	Time (s)	8.3 (3.0)	6.1 (2.5)	1.406	11	0.187	0.78 (-0.35 – 1.91)
		% time in AOI§	53.3 (27.6)	72.9 (13.1)	-1.672	8.8	0.130	-0.93 (-2.08 – 0.22)
	2 (n = 10)	Time (s)	5.3 (3.2)	6.1 (2.5)	-0.537	14	0.600	-0.28 (-1.29 – 0.74)
		% time in AOI	58.5 (14.7)	72.9 (13.1)	-1.968	14	0.069	-1.02 (-2.09 – 0.06)
	3 (n = 9)	Time (s)	3.8 (3.0)	6.1 (2.5)	-1.614	13	0.131	-0.85 (-1.93 – 0.23)
		% time in AOI	62.6 (27.6)	72.9 (13.1)	-0.845	13	0.413	-0.45 (-1.49 – 0.60)
MCT	1 (n = 7)	Time (s)§	5.9 (3.3)	1.6 (1.0)	3.252	7.4	0.013*	1.81 (0.52 –

(Image 9)								3.10)
		% time in AOI	45.5 (21.5)	73.3 (31.8)	-1.870	11	0.088	-1.04 (-2.20 – 0.12)
	2 (*n* = 10)	Time (s)	3.7 (1.5)	1.6 (1.0)	2.880	14	0.012*	1.49 (0.35 – 2.62)
		% time in AOI	49.4 (16.5)	73.3 (31.8)	-1.998	14	0.066	-1.03 (-2.11 – 0.04)
	3 (*n* = 9)	Time (s)	2.5 (0.8)	1.6 (1.0)	1.730	13	0.107	0.91 (-0.17 – 2.00)
		% time in AOI	51.3 (28.2)	73.3 (31.8)	-1.408	13	0.183	-0.74 (-1.81 – 0.32)
PCT (Image 7)	1 (*n* = 8)	Time (s)	17.6 (15.1)	5.7 (2.9)	1.885	12	0.084	1.02 (-0.11 – 2.14)
		% time in AOI§	45.2 (4.5)	60.8 (9.8)	-3.617	6.6	0.010*	-1.95 (-3.23 - -0.67)
	2 (*n* = 10)	Time (s)§	22.4 (27.3)	5.7 (2.9)	1.477	14	0.162	0.76 (-0.28 – 1.81)
		% time in AOI	47.9 (14.7)	60.8 (9.8)	-1.875	14	0.082	-0.97 (-2.03 – 0.10)
	3 (*n* = 9)	Time (s)	11.3 (11.3)	5.7 (2.9)	1.170	13	0.263	0.62 (-0.44 – 1.67)
		% time in AOI	49.9 (17.8)	60.8 (9.8)	-1.351	13	0.200	-0.71 (-1.78 – 0.35)

*$p < 0.017$. § Levene's Test for Equality of Variances not equal therefore equal variances not assumed. ‡ Cohen's $d > 0.80$ large effect size difference.

Table 4.38 Chi-square test comparing the diagnostic accuracy (percent correct) of post-test teaching intervention Groups 1, 2 and 3 and all groups combined (1-3) to the corresponding results for the expert group (Objective 1) (n = 6) for the five diagnoses (cutaneous histiocytoma [CH] (image 6), cutaneous lymphoma [LSA] (image 10), transmissible venereal tumor [TVT] (image 8), mast cell tumor [MCT] (image 9) and extramedullary plasmacytoma [PCT] (image 7).

Diagnosis	Teaching intervention Group	Post-test % correct	Expert % correct	*df*	χ^2	Sig. *p*
Cutaneous histiocytoma (Image 6)	1 (n = 8)	62.5	83.3	1	0.729	0.393
	2 (n = 10)	60.0	83.3	1	0.950	0.330
	3 (n = 10)	80.0	83.3	1	0.027	0.869
Cutaneous lymphoma (Image 10)	1 (n = 8)	87.5	83.3	1	0.049	0.825
	2 (n = 10)	100.0	83.3	1	1.778	0.182
	3 (n = 10)	100.0	83.3	-	1.778	0.182
Transmissible venereal tumor (Image 8)	1 (n = 8)	87.5	100.0	1	0.808	0.369
	2 (n = 10) §	100.0	100.0	-	-	-
	3 (n = 10) §	100.0	100.0	-	-	-
Mast cell tumor (Image 9)	1 (n = 8) §	100.0	100.0	-	-	-
	2 (n = 10) §	100.0	100.0	-	-	-
	3 (n = 10) §	100.0	100.0	-	-	-
Extramedullary plasmacytoma (Image 7)	1 (n = 8)	75.0	100.0	1	1.750	0.186
	2 (n = 10)	90.0	100.0	1	0.640	0.424
	3 (n = 10) §	100.0	100.0	-	-	-

§ Chi-square test was not performed for this group as the post-test and expert percentage correct are constant.

4.3.5.1.2 Aggregated results post-test diagnoses

For the dependent variables time to diagnosis (s) and percentage time spent in the AOIs, independent sample *t*-tests were used to compare the aggregated results of the post-test diagnoses for teaching intervention Groups one to three with the aggregated results of the experts for images 6-10 (Table 4.39). The aggregated results for diagnostic accuracy the post-test diagnoses for teaching intervention Groups one to three were compared to the aggregated results of the experts (images 1-6) using a chi-square test (Table 4.40).

For time to diagnosis (s), there was no significant difference between experts and traditional teaching intervention Group one ($p > 0.017$) and experts and basic visual teaching intervention Group two ($p > 0.017$) but a significant difference between the means of experts (M = 16.6s, SD = 22.6s) and extended visual teaching intervention Group three (M = 8.5s, SD = 9.8s) $t(85.1) = -2.483$, $p < 0.017$. There was no significant difference between experts and extended visual teaching intervention Group three for percentage time spent in the AOIs ($p > 0.017$). However, traditional teaching intervention Group one (M = 49.0%, SD = 19.7%) spent on average less percentage time in the AOIs than the experts (M = 60.0%, SD = 19.6%) $t(94) = -2.665$, $p < 0.017$. Similarly, basic visual teaching intervention Group two (M = 50.0%, SD = 18.6%) spent on average less percentage time in the AOIs than the experts (M = 60.0%, SD = 19.6%) $t(108) = -2.724$, $p < 0.017$ (Table 4.39). There was no significant difference between teaching intervention groups one, two and three and the experts for diagnostic accuracy ($p > 0.017$) (Table 4.40).

Table 4.39 Independent sample *t*-test comparing post-test teaching intervention Groups 1, 2 and 3 to the corresponding results from the expert group (n = 60) (Objective 1) for time to diagnosis (s) and percentage time spent in the AOIs (%) for all diagnoses combined.

Teaching intervention group	Slide *n*	Parameter	Post-test aggregated *M* (*SD*)	Expert aggregated *M* (*SD*	*t*-value	*df*	*p*-value	Effect size Cohen's *d* ‡ (95% CI)
1 (n = 7)	36	Time (s)§	14.9 (14.6)	16.6 (22.6)	-0.493	93.4	0.664	-0.10 (-0.52 – 0.31)
		% time in AOI	49.0 (19.7)	60.0 (19.6)	-2.655	94	0.009*	-0.56 (-0.98 - -0.14)
2 (n = 10)	50	Time (s)§	12.0 (14.7)	16.6 (22.6)	-1.281	102.4	0.203	-0.15 (-0.62 – 0.13)
		% time in AOI	50.0 (18.6)	60.0 (19.6)	-2.724	108	0.008*	-0.52 (-0.90 - -0.14)
3 (n=9)	45	Time (s)§	8.5 (9.8)	16.6 (22.6)	-2.483	85.1	0.015*	0.49 (-0.88 - -0.10)
		% time in AOI	53.3 (25.2)	60.0 (19.6)	-1.545	103	0.126	-0.30 (-0.69 – 0.08)

*p < 0.017. § Levene's Test for Equality of Variances not equal therefore equal variances not assumed. ‡ Effect size difference (Cohen's *d):* small = 0.30 to 0.49, medium = 0.50 to 0.79, > 0.80 large.

Table 4.40 Chi-square test comparing post-test teaching intervention Groups 1, 2 and 3 to the corresponding results from the expert group (n = 60) (Objective 1) for diagnostic accuracy (percent correct) for all diagnoses combined.

Teaching intervention Group	*n*	Post-test aggregated % correct	Expert aggregated % correct	*df*	χ^2	Sig. *p*
1	100	82.5	93.3	1	2.877	0.090
2	100	90.0	93.3	1	0.403	0.525
3	100	96.0	93.3	1	1.566	0.211

Chapter Five: **Discussion**

5.1 Eye-tracking technology in visual reasoning

Although eye-tracking has been used in studies of visual reasoning in radiology[72, 74-76, 78], and a limited number of studies in human pathology[6, 7], this is the first time that eye-tracking technology has been used to assess novice and expert differences in veterinary pathology. Using the baseline differences between novices and experts as a framework for the investigation of visual diagnostic reasoning in veterinary pathology, I also used eye tracking to assess two different visual diagnostic reasoning teaching interventions with DVM students at the University of Calgary.

Using tracked eye movements, I was able to demonstrate quantifiable and qualitative differences between the novice DVM students and expert veterinary pathologists. Our findings were similar to what has been published for eye-tracking assessment of visual reasoning skills in radiology: 1) experts spent less total time analyzing the image, 2) had fewer focal points of interest, and 3) fixated longer on diagnostically useful points of interest.[74-76] This would support the construct validity of the eye-tracking technology used to assess knowledge and skill development in that the experts consistently outperformed the novices. Similarly, the movement of novices toward expert eye-tracking behaviours in Objective 2 (i.e., reduced time to diagnosis, more percentage time spent in the AOIs, more efficient eye-movement patterns) would support the construct validity of eye-tracking technology as an assessment method to understanding how best to enhance the teaching and learning environment that is largely based on learners' visual diagnostic reasoning abilities.

The eye-tracking data from this study complimented the think aloud protocol by allowing for detailed and accurate measures of participants' eye-movement behaviours to be connected in

real time to their visual diagnostic reasoning process. Concurrent validity was achieved in that similar data related to the images' key features was both viewed and identified simultaneously from the eye-tracking and the think aloud protocol, respectively. In particular, this included 1) specific points of diagnostic interest identified *a priori* by the experts (i.e., AOIs [eye-tracking] and key diagnostic features verbalized [think aloud]); 2) shorter viewing time by the experts (eye-tracking) and an often quick verbalization of the diagnosis (think aloud); and 3) more efficient patterns of eye-movement (eye-tracking) correlating to more succinct verbal summation of diagnostic features by the experts (i.e., experts used higher level interpretations in their think aloud and rarely mentioned non-diagnostic or irrelevant features).

Running a concurrent think-aloud protocol with the eye tracking did, as could be expected, alter some of the eye-tracking data. In Objective 1, for all participants as well as the novice and expert groups separately, participants took significantly longer to reach a diagnosis when presented with the think aloud with eye-tracking (TA) slides (images 1 to 5) compared to the eye-tracking only (ET) slides (images 6 to 10). However, the percentage time spent visualizing the AOIs was no different between TA and ET slides for all participants as well as for the novice and expert groups separately. There were qualitative differences in the eye-tracking patterns in experts only, where patterns produced during visualization of the TA slides were more complex compared to the ET slides. A clear pattern of eye-movement between the AOIs, however, was still discernable for experts on the TA slides. This suggests that the qualitative differences in eye-movement pattern in experts noted between the TA and ET slides could be a function of the longer time the TA protocol took. In part, the experts efficiency in speed and accuracy of visual diagnostic reasoning with the ET slides reflects pattern recognition (System 1: non-analytic reasoning) and a shift to a TA protocol may force experts to slow down

the process and articulate what key features lead them to the same diagnoses (System 2: analytic reasoning). There was no discernable difference between the time that experts spent visualizing the AOIs in the TA and ET slides. Together, these data suggest that eye tracking alone may actually remove some of the consequential validity (increased time associated with talking and increased analytical thinking) that hampers visual reasoning assessment using think-aloud protocols.

5.2 Novice and Expert study

In Objective 1, I used eye-tracking data and think-aloud protocols to establish baseline differences between novice and expert pathologists. By running concurrent and retrospective think aloud protocols with the eye-tracking, I further gained insight into the reasoning modalities used by novices and experts, particularly with respect to dual-process theory and illness scripts. Using the time to diagnosis, diagnostic accuracy, percentage time viewers spent visualizing areas of interest, verbally-identified key diagnostic features, reasoning processes verbalized in think-aloud protocols, and the qualitative patterns of eye-movement, I demonstrated several differences between novice and expert veterinary pathologists. Further, our data supports our hypotheses that experts use more System-1 reasoning processes than novices and that they possess "visual illness scripts", similar to other disciplines.[54, 55]

5.2.1 Experts were quicker to formulate a diagnosis and had higher diagnostic accuracy

In seven of ten images in the study, as well as the images aggregated together, experts provided a diagnosis more quickly than novices resulting in a large effect size difference ($d >$ 1.00). Experts had a significantly higher rate of diagnostic accuracy with an overall percentage

difference from novices of 81.5%. It is not at all surprising that experts were significantly quicker to reach a diagnosis as this is consistent with other novice and expert studies in visual reasoning.[6, 69] Similarly, the mere statement of expertise suggests an expected higher level of diagnostic efficiency (i.e., time to diagnosis) and accuracy (i.e., most relevant key features and diagnosis identified correctly). The quickness of experts to reach a diagnosis can be explained in several ways.

System 1 modes of reasoning are characterized by fast, "intuitive" reasoning, heavily reliant on pattern recognition.[1] The speed that experts in this study (M = 16.6s, SD = 22.6s) reached a diagnosis would be supportive of a heavier reliance by experts on System 1 over System 2 reasoning processes. In fact, one expert reached a diagnosis within 0.1 seconds of viewing the image. It is hard to rationalize that any reasoning process other than pattern recognition (i.e., System 1 reasoning) is at play with a time to diagnosis as short as this. The data from the think aloud protocols support this observation. Experts quickly (often within a few seconds of viewing the image) stated a diagnostic hypothesis, or a short list of refined differential diagnoses. This suggests that, at least initially, experts are highly reliant on System 1 reasoning to formulate their diagnostic hypothesis(es). Time has been shown in other studies to be an indicator of when individuals are using System 1 and System 2 reasoning processes. Recent studies in the role of System 1 and System 2 cognition in clinical reasoning have used time pressure to force participants into a System 1 mode of reasoning, and time to diagnosis was used to categorize non-analytic (System 1) thinking.[61, 113] An initial apparent reliance of experts on System 1 (pattern recognition) reasoning would narrow the differentials down quickly, reducing their time to diagnosis and compromise accuracy if novel or unrecognized clinical information is presented. A related clinical reasoning theory, script theory where knowledge structures of key

features or signs and symptoms guide the individual to the correct diagnosis, could also explain the short time experts take to reach a diagnostic hypothesis after viewing an image.

In other domains, experts have been shown to have more complex and organized knowledge structures.[29] This organization is thought to allow experts to access their knowledge expediently and thus take less time to reach a diagnosis. In this manner, experts, after perceiving and interpreting the features of the case, activate a limited repertoire of illness scripts (differential diagnoses).[35, 36] Illness scripts have hierarchical knowledge structures "slots" containing relevant medical information pertaining to the clinical signs, including pre-stored knowledge about different predisposing conditions, pathophysiology and diagnostic testing, acceptable or not acceptable values and default values.[34-36] This rapid activation of knowledge, explains, in part, the quickness of an expert to formulate a diagnosis, and is likely occurring in our experts as well. In visual reasoning, it is likely that similar schemes exist that includes visual key features of the slide images that are critical to the correct diagnosis. This is supported by our data where experts spend more percentage time in the AOIs and quickly identify a greater numbers of quality key diagnostic features (see section 5.2.2).

In other disciplines, experts have been shown to have much more efficient patterns of eye-movement and that this efficiency reduces the amount of time they need to visualize the image in high foveal resolution, reducing their time to diagnosis.[69, 71, 72] Our experts also demonstrated more efficient patterns of eye-movement (see section 5.2.3), a further explanation for the shorter time experts took to reach a diagnosis.

5.2.2 Experts spent more percentage viewing time in the areas of diagnostic interest (AOIs) and identified more key diagnostic features

Experts spent significantly more time in the areas of diagnostic interest (AOIs) than novices with a large effect size difference ($d > 1.00$). In the retrospective and concurrent think-aloud protocols, experts identified more key diagnostic features with greater frequency on all of the slide images than the novices and these key diagnostic features corresponded to the diagnostic features selected *a priori* for the AOIs. This finding provides evidence for the content and construct validity of the identified AOIs for each of slide images.

Novices, in contrast, observed fewer key diagnostic features and spent significantly less time viewing the AOIs (i.e., viewed outside the identified AOI areas). They tended to make a diagnosis (even when correct) based on only one verbally identified key diagnostic feature. Novices also had a difficult time coming up with more than one differential diagnosis and had a limited repertoire of visual diagnostic features to identify from each of the slide images. This would suggest that novices had a less comprehensive or elaborate knowledge structure and, therefore, a more limited visual repertoire of exemplars to recall from than experts.

Novices identified more non-diagnostic or irrelevant features than experts. For example, novices, on almost every slide, mentioned the presence of erythrocytes (red blood cells). Although experts appeared to look at the red blood cells (based on the patterns of eye-tracking movement), they never mentioned their presence in the think aloud protocols. The presence of red blood cells in all of the images was evidence of blood-contamination, a feature that was not helpful in formulating a diagnosis. Novices also more frequently commented on variations in the slide staining, whereas experts never did. Although this was a prominent visual characteristic of one of the slide images, it was an irrelevant non-diagnostic feature. According to theories related

to the concept of using illness scripts, experts have compiled elaborate knowledge networks and are better able to distinguish diagnostically useful data allowing them to focus only on the aspects of the problem critical to the diagnosis while ignoring or filtering out irrelevant information.[30, 114] This may explain why experts in our study didn't mention non-diagnostic or irrelevant features as they quickly dismissed them nonverbally from the problem-solving or diagnostic reasoning process.

Novices tended to describe features of the slide in basic syntactic elements (i.e., colour, shape, size, etc.) rather than using more interpretive, higher level statements. For example, a novice described a lymphoblast as "the cells are large and haven't got much of an envelope (cytoplasm) around them" (novice 13, image 2). Describing the same lymphoblasts, expert 4 stated "the cells are lymphoblasts and based on the presence of lymphoglandular bodies I would bet this is a B-cell lymphoma". This would be consistent with the expert organizing the information into a semantically meaningful rationale or statement; a feature of expertise in other domains.[114] Similarly, novices provided detailed descriptions of mitotic figures "those two cells might actually be dividing because there's two cytoplasm and maybe two nuclei" (Novice 30, image 4). Schmidt et al.[33, 37] (1990, 2007) identified four levels of illness script development: 1) elaborated causal networks linking concepts based on cause and consequence, 2) abridged networks, 3) illness scripts, and 4) instance scripts. The novice's reliance on simple descriptors suggests an attempt to understand an image based on simple or basic components linked to cause and effect (elaborated). With experience, these basic features amalgamate into abridged networks (e.g., "mitotic figure") and with further experience an illness script is expressed by an expert as a "mitotic figure - this is likely a high-grade mast cell tumor" (expert 4, image 4).

As this was a preliminary study, individual images were analyzed for the percentage time spent within the AOIs by participants. There were several of the images (images 1 [cutaneous histiocytoma, CH], 2 [cutaneous lymphoma, LSA], 3 [transmissible venereal tumor, TVT], 5 [extramedullary plasmacytoma, PCT], 8[TVT], 9 [MCT] and 10 [LSA]), where there was no significant difference for the individual images found between novices and experts for the percentage time spent in the AOIs. One of the images (1 CH), though diagnostically representative, contained predominantly cells of diagnostic interest, thus the majority of nucleated cells on the image were included in the AOIs. It is then reasonable to suppose that these cells would attract visual attention regardless of the viewer's understanding of their significance. Four of the images LSA (2), PCT (5) and TVT (3 and 8) were highly cellular with most of the cells in the image showing key diagnostic features. Although AOIs were drawn on cells that were thought to be the most representative, the diagnosis could be derived from cells not included in the AOI. Prioritizing AOIs, and including more non-diagnostic cells in the image that were visually similar to the diagnostic cells would allow for these images to be more discriminating in future studies.

There were discrepancies between the paired diagnoses for key diagnostic features identified by the experts. Some of the identified features that were flagged in one image were not present in the other (for example, infiltrating lymphocytes were present in the CH image 1 but not 6). Most key features, however, were present in both slides, however occasionally these key features were mentioned in one of the paired image by experts (e.g., pale blue cytoplasm in CH) but not in the other. This discrepancy in verbalized responses from the expert group could not be explained. A few of the key diagnostic features (e.g., cytoplasmic vacuoles in the TVT and purple granules in the mast cell tumor [MCT] images) were so visually prominent that there

was no difference between novices and experts in their frequency of observing them, however the interpretation of these features between novices and experts differed – experts almost always identified them as a key identifying feature of a TVT or MCT while most of the novices could not explain their presence (particularly for the TVT).

5.2.3 Experts had more efficient patterns of eye-movement than novices

For each participant, a compilation of the eye-movement pattern for an image was taken at the end of each viewing session. Experts had more focused patterns of eye-tracking movement, and quickly identified and moved deliberately between the AOIs. In some cases, experts stated a diagnosis before discernable eye-movement was recorded. On the other hand, novices had highly irregular and seemingly erratic patterns of eye-movement with greater numbers of fixations and saccades. The efficiency of eye-movement patterns in experts is supported by the shorter time that experts took to view and diagnose an image, as well as their increased attention within the AOIs. Our data is consistent with findings in human radiology where experts initially view an image globally and holistically before rapidly formulating a diagnostic hypothesis,[71] and then seek visual cues to confirm of dispute their diagnostic hypothesis.[72] In an eye-tracking study of human pathologists, there were similar observations reported where experts quickly identified 'zoom locations' (or areas of diagnostic interest) through an initial global impression and peripheral vision. This is thought to reduce the time needed by pathologists to examine the entirety of the image in foveal vision and, thereby increasing their diagnostic efficiency.[69]

It could be speculated, that a succinct and targeted pattern of eye-movement in experts is also supportive of more efficient visual diagnostic reasoning processes. Non-analytical (System

1) reasoning is holistic in that the global impression of the case is compared to prior examples or exemplars.[39] Experts coming to a diagnosis with minimal viewing of the image would support a global impression and System 1 reasoning. System 2 (analytical) reasoning analyzes a case on a more systematic, feature-by-feature basis.[39] An analytical (System 2) mode of reasoning could be inferred from the complex pattern of eye-movement in novices with greater numbers of fixations and saccades.

Literature from cognitive psychology seems to support our speculation that eye-movement patterns are, in part, a product of the visual diagnostic reasoning process. Grant and Spivey[115] (2003) demonstrated that particular eye-movement patterns were associated with higher success in solving diagram-based (visual) problems, and that these eye-movement patterns were targeted to visual features critical to solving the problem. These eye-movement patterns, further, could be influenced by pointing out visual features critical to solving the problem (e.g., akin to pointing out key diagnostic features in our studies) and this increased the success in future problem-solving.[115] Tomas and Lleras[116] (2007) found that explicitly directing the eye-movement pattern of participants (not just pointing out critical visual features) further improved problem solving success in visual problems.

5.2.4 Experts employed pattern-recognition (system 1) and script-inductive reasoning with analytic (system 2) justification of the diagnosis

Together, the verbal and eye-tracking data from our study supports an efficient and highly organize pattern of eye-movement and reasoning in our experts, where experts would move quickly to a stated diagnosis and then spend the remainder of the viewing time justifying their diagnosis verbally (in the think-aloud protocol) and visualizing the AOIs (eye-tracking

data). Similar eye-tracking data has been shown in human radiology studies, and has lead to the suggestion that the pattern of eye-movement demonstrated by experts is consistent with higher levels of reasoning and knowledge organization.[69] Further, this pattern could also be described as an initial System 1 (pattern recognition) mode of thinking with a subsequent System 2 (analytical) mode of thinking to support or justify the diagnosis.

Novices, tended to demonstrate a predominantly analytical (System 2) approach to the diagnosis based on both their eye-tracking data (i.e., a longer time to reach a diagnosis, more irregular visual eye-movement patterns) and think aloud data (i.e., novices stated specific visual features using basic semantic elements, attempted to interpret each finding and later collated the information to come up with a diagnosis). However, the novices would at times attempt pattern-recognizing (System 1) behaviours, regardless of whether their mental prototype was correct. For example, "I am hoping for some inspiration to trigger a memory of what I am looking at" (novice 30); "This really reminds me of something. It is very familiar. I think the uniform nuclei means they are hepatocytes?" [they were actually plasma cells] (novice 17, image 5). It has been shown that an over-reliance of novice diagnosticians on non-analytical (System 1) clinical reasoning can lead to diagnostic error, likely as a result of an inadequate repertoire of (visual) exemplars,[12] as shown by the response above for novice 17.

For both novices and experts, in cases where they were uncertain about the diagnosis, there was a tendency to spend more time viewing the image and weighing different diagnostic feature options to support one of two (or more) diagnoses: "the presence of a peri-nuclear clear zone and the color of the cytoplasm (medium-blue) makes me think that this is more likely a plasmacytoma than a histiocytoma" (expert 24, image 5). This would suggest that similar to other areas, System 2 reasoning was employed more frequently in ambiguous or uncertain

cases.[113] "Adaptive expertise" or the ability of experts to adapt their problem-solving strategies is a key competency of expertize.[117] Moving from an automated System 1 predominant mode of reasoning to an analytic System 2 reasoning pattern in the face of uncertainty is consistent with adaptive expertize.

Experts quickly identified the key diagnostic features both verbally and visually, suggesting a fast retrieval of knowledge structures associated with the diagnosis. In the talk aloud protocol, experts would mention, shortly after forming a diagnosis, what features they were "looking for" to support or disprove their diagnosis "the presence of a peri-nuclear clear zone and the colour of the cytoplasm (medium-blue) makes me think that this is more likely a plasmacytoma than a histiocytoma" (expert 24, image 5). This fast retrieval of key diagnostic features of the tumor is akin to the knowledge structures accessed quickly with the use of illness scripts. On the contrary, novices uttered findings and interpretations in a less organized manner that suggests a less organized knowledge structure. Qualitatively, the organized knowledge networks that experts possess could be reflected in the organized and efficient manner that they viewed the image (i.e., pattern of eye-movement). Similarly, novices' disorganized knowledge structures are reflected in the disorganized and prolonged eye-tracking movement patterns they produced.

5.2.5 Conclusions of the novice and expert study (Objective 1)

In this study, I sought to establish baseline differences in novice student's and expert veterinary pathologist's visual reasoning skills based on eye-tracking and think-aloud data. Further, I sought to investigate the visual diagnostic reasoning strategies of experts and novices, particularly with respect to the dual-processing theory. I demonstrated that compared to novices:

1) experts take less time to formulate a diagnosis, 2) have a higher degree of diagnostic accuracy, 3) spend a higher proportion of their viewing time viewing diagnostically useful areas of interest, 4) are able to identify more key diagnostic features, and 5) have more efficient patterns of eye-movement.

The data also supports the premise that visual diagnostic reasoning has components of both System1 and System 2 cognitive processes, with experts more reliant on System 1 reasoning as experts have more efficient and targeted viewing of a diagnostic slide image. Anecdotally, many visual diagnosticians describe what is System 1 thinking where on immediate exposure to an image a diagnosis simply "pops to mind". Our data supports our hypothesis that experts develop a sophisticated set of visual "illness scripts", where visual images attained through an array of experiences are the trigger for an elaborate knowledge structure with associated key diagnostic features, disease probabilities, and pathophysiology and anatomy and prognostic outcomes.

After establishing base-line differences with the novice and expert groups in their visual reasoning and diagnostic skills using the eye-tracking data and think aloud protocol, I was interested to see if I could use modified teaching intervention strategies to alter novice (student) visual reasoning skills, towards those demonstrated by the experts. The same 10 slide images and eye-tracking metrics gathered from the novice-expert study were used to assess the diagnostic reasoning skills of the DVM students pre- and post-intervention.

5.3 Visual reasoning educational interventions

I introduced three different educational interventions in an attempt to improve the visual reasoning skills of DVM students (novices) using eye-tracking metrics derived from the experts

from Objective 1. In Objective 1, I showed that experts had a higher degree of diagnostic accuracy, took a shorter time than novices to reach a diagnosis, and spent more of their percentage viewing time visualizing the AOIs in comparison to novices. Using these metrics as a standard for diagnostic expertise, I used the eye-tracking data to compare the performance of the students from each of the three teaching interventions: Group 1) traditional, didactic teaching; Group 2) basic visual reasoning teaching intervention with active learning activities based on single images; and Group 3) extended visual reasoning teaching intervention with active learning activities using multiple images and repetition. Our expectations were that students in Group 3, and to a lesser degree Group 2, would achieve higher diagnostic accuracy, take less time to formulate a diagnosis and would spend more of their percentage time visualizing the AOIs than students in Group 1. Parallel images of the same five subcutaneous tumors (cutaneous histiocytoma [CH], cutaneous lymphoma [LSA], transmissible venereal tumor [TVT], mast cell tumor [MCT], and extramedullary plasmacytoma [PCT]) as was used in the novice and expert study were used for the pre- and post-test assessments.

5.3.1 Quantitative and qualitative differences between the teaching intervention Groups 1, 2 and 3

5.3.1.1 There was no significant difference between the three teaching intervention groups for diagnostic accuracy in the post-test.

All three groups demonstrated improvement in their diagnostic accuracy from the pre- to post-tests. However, when just the post-test results for the three groups were compared, no significant difference between the three groups was shown for the individual tumor diagnoses or for the diagnoses aggregated together.

Although no significant difference could be demonstrated between the three teaching intervention groups for diagnostic accuracy, the data may suggest some support for our hypothesis that students in teaching intervention Groups 3 and 2 would show higher diagnostic accuracy post-test. When all the diagnoses were aggregated together, the aggregated diagnostic accuracy for Group 1 was 82.5%, Group 2 was 90.0% and Group 3 was 96.0%. In four of the five parallel diagnoses (CH, LSA, TVT and PCT), Group 3 and 2 students showed a significant increase in diagnostic accuracy between the pre- and post-test. Group 1 students showed a significant increase in diagnostic accuracy between the pre- and post-test for only two of the five parallel diagnoses (CH and TVT). Similarly, in four of the five parallel diagnoses, Group 3 and Group 2 students had larger percentage differences between their pre- and post-test diagnostic accuracy scores than Group 1. This was with the exception of the CH, where only Group 3 had a higher percentage difference, with Group 2 and Group 1 showing comparable percentage differences. Potentially, with a greater sample size or more intense or longer visual reasoning education intervention the use of this type of teaching strategy to enhance visual diagnostic reasoning skills may have shown statistical significance. While diagnostic accuracy measures the end point of the diagnostic reasoning process, I had anticipated that the eye-tracking data collected would demonstrate subtler differences between the three intervention groups.

5.3.1.2 Group 3 students had a significantly greater improvement in their time to diagnosis than students in Groups 1 and 2 for the individual diagnoses.

Group 3 students showed a significantly greater reduction in the time needed to reach a diagnosis compared to Group 2 and 1 students. In particular, there was a significant decrease found in the time to diagnosis between the pre- and post-tests across all parallel pre and post-test

diagnoses for Group 3. Basic visual reasoning teaching intervention Group 2 and traditional teaching intervention Group 1, showed a significant improvement in the time to reach a diagnosis for only 1 of the 5 diagnoses, the TVT. When the post-test time to diagnosis for the three teaching intervention groups were compared (using an ANCOVA and Tukey's *post hoc* test), Group 3 was found to have taken significantly less time to reach a diagnosis post-test than Group 1 for the MCT images. No significant differences were found for time to diagnosis between the three groups for the CH, LSA, TVT or PCT slide images.

When the five diagnoses were aggregated the between group differences were no longer significant. Though, similar to diagnostic accuracy, the aggregated results show some support for our hypothesis: Group 3 time to diagnosis M = 8.5 sec (2.0 sec), Group 2 time to diagnosis M = 11.9 sec (1.9 sec), and Group 1 time to diagnosis M = 15.0 sec (2.2 sec). Again, with a larger sample size or more intense or longer visual education intervention the use of multiple visual images as a teaching strategy to enhance diagnostic reasoning may have resulted in better performance outcomes.

5.3.1.3 There was no significant difference between groups for the percentage time spent visualizing the AOIs post-test

When just the post-test percentage time spent visualizing the AOIs for the three groups was compared, there was no significant difference between the three teaching intervention groups for the five diagnoses individually or as an aggregated mean total. Similar to the diagnostic accuracy and time to diagnosis, there was some support of our hypothesis that students in Groups 3 and 2 would show a greater improvement in the percentage time spent visualizing the AOIs from the pre- to post-tests as compared to Group 1. For two of the five

diagnoses (i.e., TVT and PCT) for Group 2 this was true. For the TVT and PCT diagnosis, Group 2 students spent a greater percentage of time in the AOIs in the post-test compared to the pre-test, with no significant differences found between the pre- and post tests for Groups 3 and 1.

No significant differences between the pre- and post-test percentage time spent in the AOIs was present for any of the teaching intervention groups for the CH and MCT diagnosis, and the percentage time spent in the AOI unexpectedly decreased for teaching intervention groups 2 and 3 from the pre- to post-tests for the LSA diagnosis. Variability in the AOI allocation for the parallel LSA images could explain the unexpected decrease in the percentage time spent in the AOIs between the pre- and post-tests. Considering there was a significant improvement in diagnostic accuracy for the two visual reasoning teaching intervention groups (Groups 2 and 3) and improved time to diagnosis in Group 3, it is likely that the images of the LSA chosen for the study were discriminating and that the discrepancy for the percentage time spent in and out of AOIs is related to the AOI selection. The percentage area of the image that the AOIs covered is notably larger for the pre-test LSA image (image 2) in comparison to the post-test LSA image (image 10) (Appendix 3a and 3b). This difference in the AOI designated areas and, hence, the eye-tracking data collected could account for a decrease in the amount of time participants spent visualizing the AOIs for the LSA post-test. However, even when this image was removed from the between group comparison, no significant differences between the three teaching intervention groups for percentage time spent in the AOI post-test was present.

No change in the percentage time spent in the AOIs for the MCT may be related to the high level of prior knowledge students had about this tumor type. The pre-test diagnostic accuracy for the MCT was 82.1% and all students post-test were able to correctly diagnose the tumor. Thus, the students likely already knew prior to the teaching intervention which AOIs

were useful to formulating a correct diagnosis in this case, and the high pre-test percentage time spent in the AOIs for the MCT in all three groups (59.1%) would support this.

As noted in section 5.2.2., no significant difference between novices and experts for the percentage time spent visualizing the AOIs was noted for seven of the images: 1 (CH), 2 (LSA), 3 (TVT), 5 (PCT), 8 (TVT), 9 (MCT) and 10 (LSA). It is likely that a lack of significant "improvement" in students in Objective 2, relates to AOI selection and prioritization as seen in Objective 1.

5.3.1.4 Qualitative analysis of the eye-tracking patterns showed an improvement in eye-movement efficiency post-teaching

Similar to the novice and expert study, a snap shot of the students' eye-movement pattern was taken at the end of each viewing session for both the pre- and post-test images. Qualitatively, eye-tracking patterns for all three teaching intervention groups between the pre- and post-test, became more like the patterns noted for experts in Objective 1 (i.e., smoother transitions and more attention spend on the AOIs). This was most noticeable in eye-movement patterns for students in Group 3 and less so in Groups 2 and 1, respectively.

5.3.2 Teaching intervention groups post-test compared to experts (Objective 1)

A final way I sought to demonstrate a difference between the three teaching intervention groups was to compare them to the expert eye-tracking data from Objective 1. I expected that the more effective the teaching interventions the closer students' eye-tracking movement and data would be to the experts, such that no significant difference would be discernable. Indeed, I demonstrated that of the three teaching groups, Group 3 post-test performance was most like the

experts. There was found to be no significant differences between Group 3 students at post-test and experts for percentage time spent in the AOIs. There was a significant difference found between Group 3 students at post-test and experts for time to diagnosis, but Group 3 students actually took less time to reach a diagnosis compared to the experts. Groups 1 and 2 were not significantly different from experts in their time to reach a diagnosis, though both groups spent significantly less percentage time in the AOIs than the experts. No significant difference was noted between any of the teaching groups at post-test and experts for diagnostic accuracy.

5.3.3 Group 3 (extended visual reasoning teaching intervention: active learning, image repetition) behave most like experts post-test

I have demonstrated through comparing post-test teaching intervention group data to expert eye-tracking data (Objective 1) and comparing the three groups' performance for time to diagnosis, percentage time spent visualizing the AOIs, and to a lesser degree diagnostic accuracy, that Group 3 extended visual reasoning education intervention (i.e., active learning, image repetition) performed more like experts post-test than Groups 1 (traditional, didactic teaching) and 2 (basic visual reasoning education intervention with active learning, but use of single images).

5.3.3.1 Students improved time to diagnosis could reflect a shift from System 1 to System 2 reasoning.

There was a significant improvement in the time to reach the diagnosis from the pre- to post-tests across all diagnoses with a significantly faster time to diagnosis at post-test for students in Group 3 compared to Groups 1 and 2. There was no significant difference between experts and Groups 1 and 2 for time to diagnosis at post-test, and Group 3 students at post-test

actually came to a diagnosis more quickly than the experts. Qualitatively, Group 3 eye-movement patterns at the post-test were most like experts, with simplified tracts that moved smoothly between the AOIs in an efficient manner.

Like the novice and expert study (Objective 1), this reduction in time to diagnosis and increase in the efficiency in eye-movements after the teaching interventions could suggest a transition of novices from more time consuming analytic reasoning (System 2) to more pattern-recognition reasoning (System 1). In other fields, with repetitive practice, there is a transformation from System 2 (effortful and time consuming) to System 1 (effortless and quick) reasoning processes.[1, 42, 43] Repeated and directed exposure of students to cases is also thought to be one of the more effective methods for students to acquire illness scripts[8-13] or in our experiment, visual diagnostic scripts. The students in Group 3 had a greater opportunity than students in Groups 1 and 2 to practice their visual diagnostic reasoning skills (i.e., on multiple case images), possibly accounting for their higher performance for time to diagnosis. System 1 reasoning is also thought to be based more on visual similarity than on verbal descriptions.[8] Allowing students to see multiple visual examples of the diagnosis (as was provided in the educational intervention for Group 3) would also likely encourage stronger System 1 reasoning. In other studies, encouraging non-analytic, pattern recognizing (System 1) reasoning by showing multiple images has also resulted in higher diagnostic accuracy and shorter time to diagnosis[118].

In educational studies of biomedical knowledge transference, there is increased transference of concepts if multiple examples are taught, compared to a single example, or even less when the conceptual principle only is taught.[99] In our study, Group 3 students were provided with the greatest number of examples of the tumor images and, hence, an increased transference was expected. This was seen in the more "expert-like" behaviour of students in

Group 3 for time and percentage time spent visualizing the AOIs, however, as the post-test was immediately after the teaching sessions it would have been potentially important to have seen if there was a difference in the three teaching groups for retention of this performance level over a period of time (i.e., a follow-up weeks or even months later).

Similar to our novice and expert study (Objective 1), the more efficient eye-movement patterns seen in all three education intervention groups at post-test, but most prominent in Group 3 students, could support a shift in students' visual diagnostic reasoning from System 2 (analytic) to System 1 (non-analytic) with extended visual reasoning teaching strategies.

5.3.3.2 Group 3 students post-test percentage time spent visualizing the AOIs was indistinguishable from experts.

I expected that teaching students the visual diagnostic features of the five subcutaneous masses would improve their diagnostic reasoning and, as a consequence, increase the amount of time they spent visualizing the diagnostic AOIs as shown by the experts (Objective 1). Students in Group 3 spent a similar percentage of their time visualizing the AOIs as the experts did, whereas students in Groups 1 and 2 spent less time than experts visualizing the AOIs. All three groups were initially taught didactically in the teaching session what the visual diagnostic features for each of the tumors were. For Group 1, this was only done verbally and not reinforced with active learning or visual highlighting. Students in Groups 2 and 3 were taught the visual diagnostic features verbally and were then asked to use those diagnostic features to compare and contrast practice cases using either single (Group 2) or multiple (Group 3) tumor images from prepared slides. The instructor for Groups 2 and 3 also visually pointed out these

features during the in-class discussion after students had an opportunity to diagnose the practice cases.

In multi-media learning, it has been shown that highlighting relevant features in a diagram lead to more successful learning outcomes.[108] It has also been shown that perceptually highlighting an area of the diagram critical to solving a visual problem (similar to pointing out the key diagnostic features in a cytology image) could guide attention and eye-movement patterns that dramatically improved visual diagnostic reasoning.[115] In our study, all students received information on what the diagnostic features for each case were, but only students in Groups 2 and 3 had those features pointed out in the images and were actively able to seek out these visual features during the practice cases. This could explain the more "expert-like" performance of Group 3 students in the percentage time they spent viewing the AOIs.

5.3.3.3 Optimized germane cognitive load may have improved Group 3 post-test performance

Cognitive load theory emphasizes the importance of optimizing the germane cognitive load and minimizing the intrinsic and extrinsic load during the reasoning and learning process. One way the germane load in clinical reasoning teaching can be optimized is by varying the cases (e.g., encouraging comparison between cases and identifying unifying principles).[84] This process was encouraged in Groups 2 and 3 to enhance student learning. A second way to optimize the germane load is to vary the types of cases encountered (e.g., mixed practice instead of blocked practice).[86, 96] This teaching strategy was also incorporated with Group 3 students, where multiple cases of the five diagnoses were randomly mixed during their intervention. Group 2 and 3 students also had the visual diagnostic features of the images demonstrated visually, whereas students in Group 1 were told what the visual diagnostic features were

verbally. In multimedia learning, it has been shown that visually highlighting relevant features on an image reduces the extraneous cognitive load and improves learning.[107, 108] Potentially improved visual reasoning performance of Group 3 and to a lesser extend Group 2 students could also be explained by the more optimized germane load established with the Group 3 and 2 teaching interventions.

5.3.4 Conclusions of the visual reasoning educational interventions study (Objective 2)

I hypothesized that the use of explicit features (i.e., teaching students the visual diagnostic features of the five subcutaneous masses), and the introduction of repetitive active learning (i.e., allowing students to practice their visual diagnostic reasoning on multiple cases in a mixed manner) would be more effective in transforming the students' visual diagnostic reasoning skills towards expert performance when compared to more traditional, didactic teaching modalities. I was able to demonstrate some differences between Group 3 (extended visual reasoning teaching intervention with active learning and multiple image repetition) and Groups 2 and 1, but were unable to demonstrate a difference between the Group 2 and 1 students' performance. Group 3 students were taught the visual diagnostic features of the five tumor types both visually and verbally, and used these features to diagnose multiple different images. Thus, students in Group 3 not only had the advantage of practicing their visual diagnostic reasoning skills multiple times, they also were shown a repertoire of visual exemplars of the tumors that would enhance their pattern recognition skills. These results would suggest that integrating active learning of multiple visual examples may aid in visual diagnostic reasoning skill development in DVM students.

The teaching of students in Group 2 (basic visual reasoning teaching intervention with active learning, but only single images) differed from Group 1 (traditional, didactic teaching) in that they had the visual features of the cases shown to them both visually and verbally. They were also able to actively use these visual diagnostic features to diagnose a single representative image of the five subcutaneous tumor cases. There was data collected that suggested Group 2 was performing better than Group 1 students on time to diagnosis, percentage time spent in the AOIs and the diagnostic accuracy, however, the findings were not significant and potentially reflect the need for research that looks at larger sample sizes and longer or more intense teaching sessions.

I suggest that the improved time to diagnosis in all teaching intervention groups represent, in part, a shift in the student's visual diagnostic reasoning modalities from a predominant System 2 focus to a more System 1 pattern recognition, nonanalytic reasoning process, though more eye-tracking data (including think aloud protocol data) would be necessary to support this hypothesis.

5.4 Limitations of the thesis

There are several limitations to the two studies conducted for this thesis research project. These include primarily the sample sizes (particularly for Objective 2), the intensity or duration of time allocated for the educational intervention sessions, time to diagnosis as a measure of expertise (Objectives 1 and 2), and the selection of AOIs (Objective 1 and 2).

5.4.1 Sample size

For the educational interventions study (Objective 2), the most notable limitation is the sample size. For all of the dependent variables (i.e., diagnostic accuracy, time to diagnosis, and percentage time spent visualizing the AOIs), the data suggested a higher performance for Group 3 followed by Group 2 and then Group 1 students. Therefore, it is anticipated that the subtler between group differences in this study would have benefitted from a larger sample size. The number of students enrolled in the DVM program in any one year, however, limits the sample sizes that can be accessed for this study at a single institution, and the eye-tracking equipment failure for three of the participants further reduced the data collected. For future studies, replicating the teaching interventions in two consecutive years of the DVM program or including students from another DVM program (e.g., doubling the sample size) may demonstrate more significant differences between the three educational intervention groups. Although a comparable sample size was studied for the novice and expert study (Objective 1), the effect sizes were so large that a significant difference was found.

The use of multiple slides and multiple analyses in the analysis of Objective 1 and Objective 2 also has the potential limitation of introducing type 1 error. In an attempt to counteract this, I used a Bonferroni adjusted *p*-value of 0.017 as well as a more rigorous ANCOVA analysis. While these measures may have helped reduced the risk of type 1 error, it is still a potential limitation of the study.

5.4.2 Intensity or duration of time allocated for the educational intervention sessions

One other major consideration for potentially strengthening or improving the outcomes in the education interventions study (Objective 2), would have been to either intensify or extend the

time allocated for the three different educational intervention sessions. While the time provided for all three sessions was relatively appropriate for the five diagnostic tumor types taught, the study might have benefited from using a greater number and more complex image diagnoses to learn (e.g., double the number of tumors). Alternatively, or in addition, the time allocated to the sessions could have been longer or extended to multiple sessions where each of the student groups may have attended two or more sessions instead of the single, one hour session provided in this study.

Although the three educational interventions were of the same duration (1 hour), the actual time within the teaching session dedicated to teaching the round cell tumors differed between the groups (i.e., 20 minutes for Group 1, 30 minutes for Group 2 and 40 minutes for Group 3). This could have potentially influenced the performance of these three groups post-test. One measure to counteract this would be to introduce a knowledge-based pre-test and post-test to the study, in addition to the pre- and post- skill test that was performed. This would help to ascertain if differences in knowledge levels between the three groups influenced their skill performance. This would have the added advantage of allowing us to further dissect the role of knowledge structures in the novice and expert study in visual diagnostic reasoning.

5.4.3 Time to diagnosis as a measure of expertise

Although a quick time to diagnosis was found to be a feature of expertise, the time to diagnosis was unexpectedly not predictive of diagnostic accuracy. The time to diagnosis in both in the novice and expert and educational interventions studies were also skewed. Based on a review of the video of the novice and expert and the educational intervention studies, the time to reach a diagnosis was fast when the diagnosis was known and correct, and also fast when the

participant was incorrect and had no idea of what the diagnosis was. On images where the diagnosis was uncertain or a narrow list of competing diagnoses was formulated, the time to diagnosis was much longer.

A predominant, System 1 clinical reasoning process can explain the fast time to diagnosis when the diagnosis is known and correct. Perhaps the short time to diagnosis for the incorrect answers too has a System 1 component, in that the participant quickly compares the image to their visual repertoire and if no similar image is present, the image is quickly dismissed as not being able to make a diagnosis. The longer time to diagnosis on cases that were ambiguous or a participant had difficulty deciding between competing differentials, would suggest an increased reliance on System 2 reasoning. Further, the longer time that participants took to reach a diagnosis in the talk aloud slides, compared to eye-tracking only, in the novice and expert study, would be consistent with verbalizing thoughts forcing a participant into a System 2 mode of reasoning[8], providing further support for time being a function of reasoning rather than of expertise. Thus, a short time to diagnosis, may not reflect expertise *per se*, as I suggest in this study, but just a predominance of System 1 reasoning. This too may partially explain why I had difficulty showing a significant difference in time to diagnosis between our teaching intervention groups. A more detailed retrospective think-aloud protocol may allow us to dissect this phenomenon further, particularly with the assumption that System 1 reasoning is in play when an incorrect diagnosis that is different from any of the participant's visual repertoire is made.

5.4.4 AOI selection

Although I am convinced that the AOIs were selected for the study were representative and valid, there were limitations based on the image software, as to how large or accurate the

AOIs could be created, and consequently how discriminating each of the individual AOIs were tracked when viewed by the participants. Many of the visual diagnostic features elicited by the experts in the think-aloud protocols were too small to form an AOI in their own right (e.g., nuclear chromatin patterns, cytoplasmic granules or vacuoles). Hence, the AOIs that were selected included "diagnostic cells" which would frequently contain many, if not all of the verbalized key diagnostic features. It would be interesting in future studies to have AOIs that represent specific key diagnostic features, and thus allow us to see if (as our think-aloud data suggests) the different key diagnostic feature are weighted differently in the reasoning process. This could be achieved by increasing the magnification of the image or by using a different type of diagnostic pathology image, for example a histopathology or gross specimen image.

Further, in some of the images, no significant difference was found between the novices and experts for the percentage time spent visualizing the AOIs. In one of these images (1 CH), though the AOIs were diagnostically representative, the AOIs contained almost all of the nucleated cells on the image. Thus, potentially, the cells in the AOIs attracted visual attention regardless of the viewer's understanding of their significance. In another four of the images (2 LSA, 3 TVT, 5 PCT and 8 TVT) the images were highly cellular with most of the cells in the image showing key diagnostic features. Although AOIs were drawn on cells that were thought to be the most representative, the diagnosis could be derived from cells not included in the AOI. Prioritizing AOIs, and including more non-diagnostic cells in the image that were visually similar to the diagnostic cells would allow for these images to be more discriminating.

5.5 Future directions

This was essentially a preliminary study investigating rudimentary differences between novice and expert pathologists and assessing three educational interventions. There are a number of directions both these studies can move in the future.

5.5.1 Future directions for eye-tracking methodology as a measurement tool for visual diagnostic reasoning

A challenge with using eye-tracking technology to assess visual diagnostic reasoning abilities is the literal mountain of data generated by the software. For each image, for each participant in the two studies conducted, there was up to 100 data points generated each 0.13s. Deriving a meaning to describe a highly complex process from this data becomes a logistical challenge. Anderson et al.[68] (2011) in their eye-tracking study of radiology images, characterized each fixation (gaze dwell on a 40 pixel radius for < 0.1s) with descriptors for expertise, anatomic location and image order. Using these criteria, they created a list of rules to differentiate novices from experts. Not only is this a highly laborious task that generated over 84,000 rules that were then manually excluded or included, but also the consequent rules used in this study were no different from what were derived from our AOIs. With subsequent experiments, I will attempt to derive more meaningful information from the AOIs including the preference of AOIs identified between the experts, the order that AOIs were visualized, and the potential correlations between viewing a particular AOI and diagnostic success.

Some of the information about the visual diagnostic reasoning processes used by experts and novices was better accessed from the think-aloud protocols than the eye-tracking data. However, eye tracking alone was a higher fidelity assessment of what happens during

visualization in real-time. Thus rather than replace think aloud protocols for diagnostic reasoning investigation, eye-tracking movement compliments it as an additional measure of the non-verbal aspects of the visual reasoning process. To reduce the interference of concurrent think aloud with eye tracking, a retrospective think-aloud protocol with more directed questions as to the reasoning process might be useful. This would allow us to collect the think aloud data without it interfering with the eye-tracking movements and reasoning processes used while the participant is viewing the image.

5.5.2 Future directions with novice and expert study: further dissecting the reasoning process

In my study, as in clinical reasoning in other fields, I seem to have just scratched the surface of the complex phenomenon that is visual diagnostic reasoning. This study stimulates several lines of inquiry into the reasoning process, including further dissecting the role of System 1 and System 2 reasoning modalities in visual reasoning. This includes investigating the role of System 1 and System 2 reasoning in both correct and incorrect diagnoses. Although several studies have investigated dual-process with respect to medical error,[113, 119] no studies to date have systematically investigated the diagnostic reasoning processes at play during an incorrect diagnosis. This information would be useful not only to improving the expertise of our experts and reducing medical error, but would also be informative to teaching novices.

It would also be interesting to investigate further the patterns of eye-movement and visual strategies the experts use beyond just formulating a diagnosis. Theoretically, if all the experts were only interested in formulating a diagnosis based the AOIs (as their verbal data would suggest), they would spend closer to 100% of their time visualizing the AOIs than the 45-75% of time they spent in this study. This suggests that experts are employing other visual strategies

that may include search strategies to ensure that the initial global impression was correct, analyzing non-diagnostic features to confirm they are non-diagnostic, and searching for secondary disease processes.

I would also be interested in investigating whether a switch from System 1 to System 2 reasoning correlates to physical differences in the speed of eye-movement in visual reasoning. In surgery, a "slowing down" of an automatic physical action has been shown to correspond to a switch to a more effortful mode of thinking (System 2). Similar to clinical reasoning, this occurred when surgeons were uncertain about a surgical process (e.g., they perceived anomalies in the anatomy, or the surgical plan was altered).[49] In this example, there was shown to be a physical slowing down of the surgeons' hand movements. It would be interesting to see if in visual reasoning, a switch to system 2 reasoning process manifests in changes in the speed of eye-movement, or more quantitative changes in the eye-movement patterns, in addition to an increased time viewing the image as was demonstrated in this study.

5.5.3 Future directions of the educational interventions

An obvious future direction for the educational interventions is to continue the study outlined in this thesis over a longer period of time to accrue more participants, or to increase the intensity and/ or the duration of the intervention. As the study was limited by the sample size (which in turn was dependent on the small class size of the DVM program at the University of Calgary), running the teaching interventions over a number of years may allow us to draw more conclusions as to the effectiveness of the teaching interventions.

Another obvious future direction is to ascertain the long-term retention of visual reasoning skills of the veterinary students weeks to months after the teaching interventions were

implemented. Cytology in the DVM program is taught in the second year of the program, then reinforced in the fourth year. Thus, there is a significant gap between the teaching and use of these skills. It would be interesting to see if different teaching modalities, their duration and intensity, are more conducive to long-term retention than others.

A challenge faced in teaching diagnostic pathology in a classroom setting is the difference (or lack of fidelity) between the classroom environment to the context where students are expected to use the knowledge (in a veterinary practice). With specific reference to cytology, veterinarians in practice visualize a prepared sample under a microscope and thus the image is viewed as a fluid (moving) image rather than as a static image and the diagnostician (DVM) needs to find representative areas of the sample to make their diagnosis. On the contrary, in class, cytology is often taught as single, static images of a highly representative area. Previously, the microscope has presented a physical barrier that makes analyzing the way experts view a fluid image inaccessible. With the development of tele-pathology (specifically the Aperioscope), high definition scans of the entire diagnostic slide can be viewed on large screens. I would be interested in studying how visual diagnostic reasoning between novices and experts differ when a fluid, rather than static, image is used, and if this can be transferred to the classroom setting (i.e., does teaching using fluid images enhance transference of visual reasoning skills better to an in-clinic situation).

5.5.4 Other future directions and conclusions

Eye-tracking technology presents an exciting new method for examining visual diagnostic reasoning that extends beyond pathology. Due to the nature of veterinary practice (i.e., veterinary patients have difficulty communicating their various ailments), visual

observational skills are relied on heavily in veterinary fields outside of the more obvious visual diagnostic fields. Lameness examination in horses, interpretation of herd behaviour in large animal medicine, neurological exams of small animals, interpretation of ECG readings, and assessment of surgical skills are only a few of the areas where eye-tracking could provide insight into veterinarian's thinking and performance processes.

References

[1] Evans JS. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol*. 2008;**59**:255-278.

[2] Hadford M. *Where's Waldo?* Somerville, MA Candlewick; 2007.

[3] Higgs J, Jones MA, Loftus S, Christensen N, eds. *Clinical reasoning in the health professions*. 3rd ed. Elsevier Butterworth Heinemann: Philadelphia, PA, 2008.

[4] Crowley RS, Naus GJ, Stewart J, 3rd, Friedman CP. Development of visual diagnostic expertise in pathology -- an information-processing study. *J Am Med Inform Assoc*. 2003;**10**:39-51.

[5] Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*. 2007;**242**:396-402.

[6] Krupinski EA, Tillack AA, Richter L, et al. Eye-movement study and human performance using telepathology virtual slides: implications for medical education and differences with experience. *Hum Pathol*. 2006;**37**:1543-1556.

[7] Tiersma ES, Peters AA, Mooij HA, Fleuren GJ. Visualising scanning patterns of pathologists in the grading of cervical intraepithelial neoplasia. *J Clin Pathol*. 2003;**56**:677-680.

[8] Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ*. 2007;**41**:1140-1145.

[9] Lee A, Joynt GM, Lee AK, et al. Using illness scripts to teach clinical reasoning skills to medical students. *Fam Med*. 2010;**42**:255-261.

[10] Bloch RF, Hofer D, Feller S, Hodel M. The role of strategy and redundancy in diagnostic reasoning. *BMC Med Educ*. 2003;**3**:1.

[11] Kassirer JP. Teaching clinical reasoning: case-based and coached. *Acad Med*. 2010;**85**:1118-1124.

[12] Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2004;**39**:98-106.

[13] Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;**39**:418-427.

[14] Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ*. 2003;**37**:695-703.

[15] Lesgold A, Rubinson H, Feltovich PJ, R. G, Klopfer D, Wang Y. Expertise in a complex skill: diagnosing x-ray pictures. In: Chi MTH, Glaser R, Farr MJ, eds. *The nature of expertise*. Erlbaum, Hillsdale, NJ, 1988; 311-341.

[16] Marshall R, Cartwright N, Mattick K. Teaching and learning pathology: a critical review of the English literature. *Med Educ*. 2004;**38**:302-313.

[17] Rhind SM. Education in veterinary pathology: innovations and global challenges. *J Vet Med Educ*. 2007;**34**:381-382.

[18] Kelly DF. Veterinary pathology in the United Kingdom: past, present, and future. *J Vet Med Educ*. 2007;**34**:383-389.

[19] Sharkey LC, Dial SM, Matz ME. Maximizing the diagnostic value of cytology in small animal practice. *Vet Clin North Am Small Anim Pract*. 2007;**37**:351-372, vii.

[20] Hamilton PW, van Diest PJ, Williams R, Gallagher AG. Do we see what we think we see? The complexities of morphological assessment. *J Pathol*. 2009;**218**:285-291.

[21] Pelaccia T, Tardif J, Triby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011;**16**.
[22] Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA Harvard University Press; 1978.
[23] Hopayian K. Why medicine still needs a scientific foundation: restating the hypotheticodeductive model - part two. *Br J Gen Pract*. 2004;**54**:402-403; discussion 404-405.
[24] Hopayian K. Why medicine still needs a scientific foundation: restating the hypotheticodeductive model - part one. *Br J Gen Pract*. 2004;**54**:400-401; discussion 404-405.
[25] Norman G, Trott AL, Brooks L, Kinsey M. Smith E. Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*. 1994;**6**:114-120.
[26] Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ*. 1987;**21**:86-91.
[27] Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. *Cognitive Science*. 1986;**10**:91-116.
[28] Norman GR, Brooks LR, Allen SW. Recall by expert medical practitioners and novices as a record of processing attention. *J Exp Psychol Learn Mem Cogn*. 1989;**15**:1166-1174.
[29] Boshuizen HC, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*. 1992;**16**:153-184.
[30] Schmidt HG, Boshuizen HP. On the origin of intermediate effects in clinical case recall. *Mem Cognit*. 1993;**21**:338-351.
[31] Rikers RM, Loyens S, te Winkel W, Schmidt HG, Sins PH. The role of biomedical knowledge in clinical reasoning: a lexical decision study. *Acad med*. 2005;**80**:945-949.
[32] de Bruin AB, Schmidt HG, Rikers RM. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Acad med*. 2005;**80**:765-773.
[33] Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med*. 1990;**65**:611-621.
[34] Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 2000;**75**:182-190.
[35] Charlin B, Boshuizen HP, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ*. 2007;**41**:1178-1184.
[36] Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002;**324**:729-732.
[37] Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007;**41**:1133-1139.
[38] Papa FJ, Shores JH, Meyer S. Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. *Acad med*. 1990;**65**:S21-22.
[39] Regehr G, Cline J, Norman GR, Brooks L. Effect of processing strategy on diagnostic skill in dermatology. *Acad med*. 1994;**69**:S34-36.
[40] Sloman SA. Two systems of reasoning. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, New York, 2002; 379-396.
[41] Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol*. 2003;**58**:697-720.
[42] Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;**84**:1022-1028.

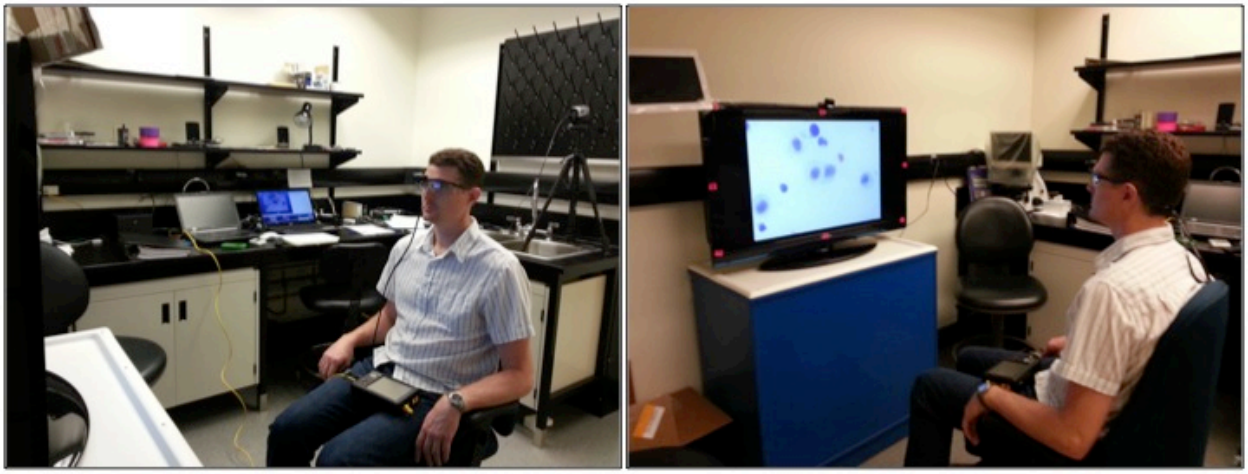
[43] Kahneman D, Frederick S. A model of heuristic judgment. In: Holyoak K, Morrison RG, eds. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, Cambridge, UK, 2005; 267-294.
[44] Norman G, Monteiro S, Sherbino J. Is clinical cognition binary or continuous? *Acad med*. 2013;**88**:1058-1060.
[45] Custers EJ. Medical education and cognitive continuum theory: an alternative perspective on medical problem solving and clinical reasoning. *Acad med*. 2013;**88**:1074-1080.
[46] Evans JS. The heuristic-analytic theory of reasoning: extension and evaluation. *Psychon Bull Rev*. 2006;**13**:378-395.
[47] Kahneman D. *Thinking Fast and Slow*. 1st ed. New York, NY Farrar, Straus and Giroux; 2011.
[48] Norman GR, Trott AL, Brooks LR, Smith EKM. Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine*. 1994;**6**:114-120.
[49] Moulton CA, Regehr G, Lingard L, Merritt C, MacRae H. Slowing down to stay out of trouble in the operating room: remaining attentive in automaticity. *Acad med*. 2010;**85**:1571-1577.
[50] Marcum JA. An integrated model of clinical reasoning: dual-process theory of cognition and metacognition. *J Eval Clin Pract*. 2012;**18**:954-961.
[51] Durning S, Artino AR, Jr., Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;**45**:927-938.
[52] Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Advances in health sciences education : theory and practice*. 2011.
[53] Engel PJH. Tacit knowledge and visual expertise in medical diagnostic reasoning: implications for medical education. *Medical Teacher*. 2008;**30**:184-188.
[54] Norman GR, Coblentz CL, Brooks LR, Babcook CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med*. 1992;**67**:S78-83.
[55] Duchowski A. Visual attention. In: Duchowski A, ed. *Eye tracking methodology: theory and practice*. 2nd. edn. SpringerLINK eBooks, 2007.
[56] Wood BP. Visual expertise. *Radiology*. 1999;**211**:1-3.
[57] Arocha JF, Patel VL. Methods in the study of clinical reasoning. In: Higgs J, Jones MA, Loftus S, Christensen N, eds. *Clinical reasoning in the health professions*. 3rd edn. Elsevier, Philadelphia, PA, 2008; 193-211.
[58] Woods NN. Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ*. 2007;**41**:1173-1177.
[59] Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen*. 1991;**120**:278-287.
[60] Kulatunga-Moruzi C, Brooks LR, Norman GR. Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine*. 2001;**13**:110-116.
[61] Mamede S, van Gog T, van den Berge K, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Jama*. 2010;**304**:1198-1203.

[62] Coderre S, Wright B, McLaughlin K. To think is good: querying an initial hypothesis reduces diagnostic error in medical students. *Acad med*. 2010;**85**:1125-1129.
[63] Arocha JF, Wang D, Patel VL. Identifying reasoning strategies in medical decision making: a methodological guide. *J Biomed Inform*. 2005;**38**:154-171.
[64] Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teaching and Learning in Medicine*. 2000;**12**:189-195.
[65] McLaughlin K, Coderre S, Mortis G, Fick G, Mandin H. Can concept sorting provide a reliable, valid and sensitive measure of medical knowledge structure? *Advances in health sciences education : theory and practice*. 2007;**12**:265-278.
[66] Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. *Acad med*. 1997;**72**:S37-39.
[67] Bordage G. Prototypes and semantic qualifiers: from past to present. *Med Educ*. 2007;**41**:1117-1121.
[68] Anderson B, Shyu CR. Studying visual behaviors from multiple eye tracking features across levels of information representation. *AMIA Annu Symp Proc*. 2011;**2011**:72-79.
[69] Krupinski EA. Current perspectives in medical image perception. *Atten Percept Psychophys*. 2010;**72**:1205-1217.
[70] Hoffman JE. Visual attention and eye movements. In: Pashler H, ed. *Attention*. Cambridge University Press, Hove, England, 1998; 119-153.
[71] Nodine CF, Kundel HL. Using eye movements to study visual search and to improve tumor detection. *Radiographics*. 1987;**7**:1241-1250.
[72] Mello-Thoms C, Hardesty L, Sumkin J, et al. Effects of lesion conspicuity on visual search in mammogram reading. *Acad Radiol*. 2005;**12**:830-840.
[73] Jarodzka H, Scheiter K, Gerjets P, van Gog T. In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*. 2010;**20**:146-154.
[74] Mello-Thoms C. Perception of breast cancer: eye-position analysis of mammogram interpretation. *Acad Radiol*. 2003;**10**:4-12.
[75] Mello-Thoms C, Dunn SM, Nodine CF, Kundel HL. The perception of breast cancers--a spatial frequency analysis of what differentiates missed from reported cancers. *IEEE Trans Med Imaging*. 2003;**22**:1297-1306.
[76] Mello-Thoms C. The problem of image interpretation in mammography: effects of lesion conspicuity on the visual search strategy of radiologists. *Br J Radiol*. 2006;**79 Spec No 2**:S111-116.
[77] Krupinski EA, Nodine CF, Kundel HL. Enhancing recognition of lesions in radiographic images using perceptual feedback. *Optical Engineering*. 1998;**37**:813-818.
[78] Mello-Thoms C. How does the perception of a lesion influence visual search strategy in mammogram reading? *Acad Radiol*. 2006;**13**:275-288.
[79] Sowden PT, Davies IR, Roling P. Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *J Exp Psychol Hum Percept Perform*. 2000;**26**:379-390.
[80] Chi MTH, Glaser R, Rees E. Expertise in problem solving. In: Sternberg RJ, ed. *Advances in the psychology of human intelligence*. Erlbaum, Hillsdale, NJ, 1982; 7-75.

[81] Voss JH, Greene TR, Post TA, Penner BC. Problem-solving skill in the social sciences. In: Bower GH, ed. *The psychology of learning and motivation: advances in research and theory*. Academic Press, New York, NY, 1983; 165-213.
[82] Lesgold A. Acquiring expertise. In: Anderson JR, Kosslyn SM, eds. *Tutorials in learning and memory: essays in honor of Gordon Bower*. Calif: Freeman, San Francisco, CA, 1984; 31-60.
[83] Merriam SB, Caffarella RS. The learning process. In: Merriam SB, Caffarella RS, eds. *Learning in adulthood: a comprehensive guide*. 2nd edn. Jossey-Bass, San Franscisco, CA, 1999; 193-220.
[84] Dale VH, Sullivan M, May SA. Adult learning in veterinary education: theory to practice. *J Vet Med Educ*. 2008;**35**:581-588.
[85] Khalil MK, Paas F, Johnson TE, Payer AF. Interactive and dynamic visualizations in teaching and learning of anatomy: a cognitive load perspective. *Anat Rec B New Anat*. 2005;**286**:8-14.
[86] van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;**44**:85-93.
[87] Mandin H, Jones A, Woloschuk W, Harasym P. Helping students learn to think like experts when solving clinical problems. *Acad Med*. 1997;**72**:173-179.
[88] Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;**355**:2217-2225.
[89] Horii CV. Teaching insights from adult learning theory. *J Vet Med Educ*. 2007;**34**:369-376.
[90] Woods NN, Brooks LR, Norman GR. The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Med Educ*. 2005;**39**:107-112.
[91] Woods NN, Neville AJ, Levinson AJ, Howey EH, Oczkowski WJ, Norman GR. The value of basic science in clinical diagnosis. *Acad med*. 2006;**81**:S124-127.
[92] Woods NN, Brooks LR, Norman GR. The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv Health Sci Educ Theory Pract*. 2007;**12**:417-426.
[93] Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educational Researcher*. 1989;**18**:32-42.
[94] Hines SA, Collins PL, Quitadamo IJ, Brahler CJ, Knudson CD, Crouch GJ. ATLes: the strategic application of Web-based technology to address learning objectives and enhance classroom discussion in a veterinary pathology course. *J Vet Med Educ*. 2005;**32**:103-112.
[95] Rosen KR, McBride JM, Drake RL. The use of simulation in medical education to enhance students' understanding of basic sciences. *Medical Teacher*. 2009;**31**:842-846.
[96] Norman G. Teaching basic science to optimize transfer. *Medical Teacher*. 2009;**31**:807-811.
[97] Bender HS, Lockee BB, Danielson JA, et al. Mechanism-based diagnostic reasoning: thoughts on teaching introductory clinical pathology. *Vet Clin Pathol*. 2000;**29**:77-83.
[98] Quilici J, Mayer RE. Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*. 1996;**88**:144-161.
[99] Catrambone R, Holyoak KJ. Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1989;**15**:1147-1156.

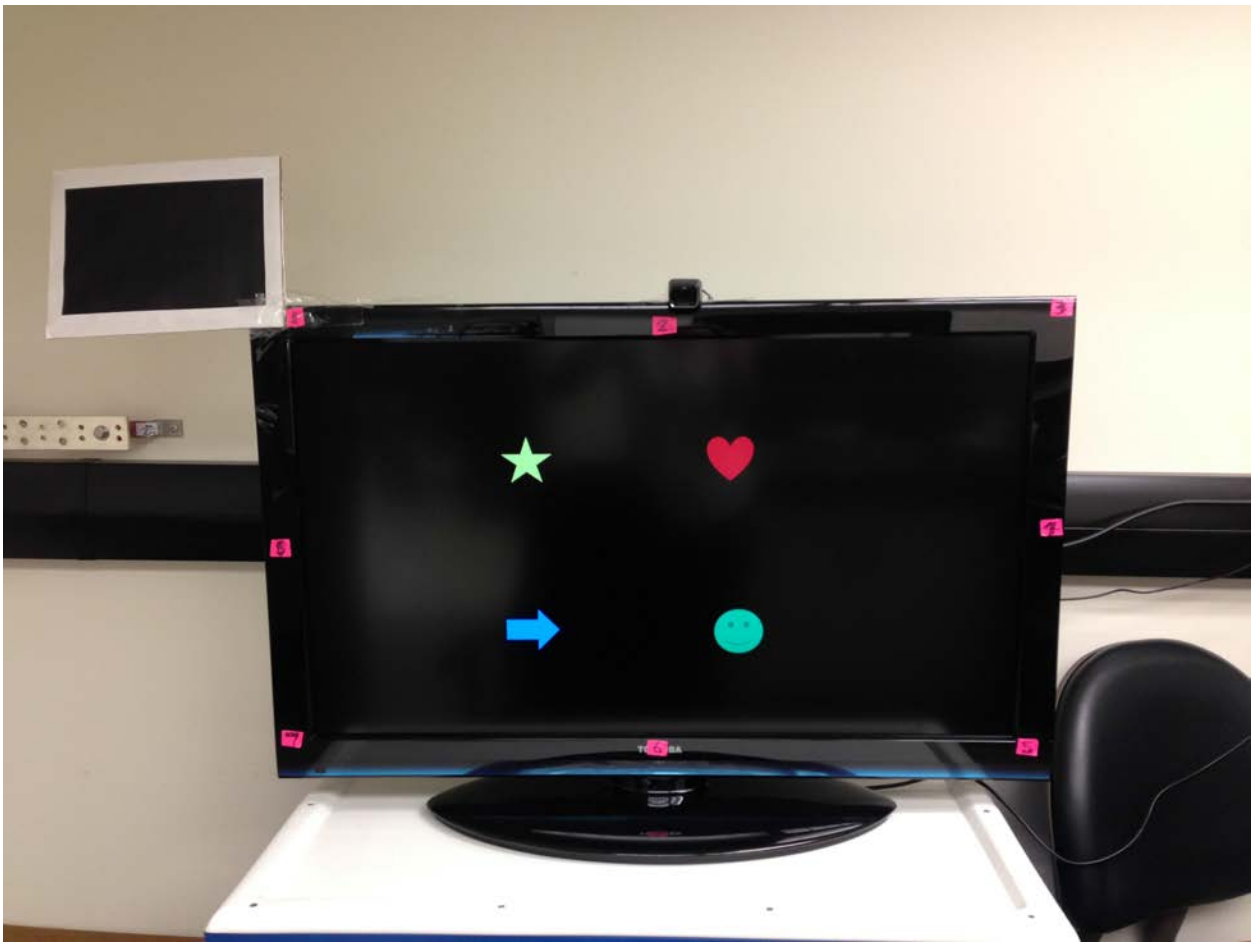
[100] Patel VL, Arocha JF, Branch T, Karlin DR. Relationship between small group problem-solving activity and lectures in health science curricula. *J Dent Educ*. 2004;**68**:1058-1080.
[101] Mandin H, Harasym P, Eagle C, Watanabe M. Developing a "clinical presentation" curriculum at the University of Calgary. *Acad Med*. 1995;**70**:186-193.
[102] Papa FJ, Harasym PH. Medical curriculum reform in North America, 1765 to the present: a cognitive science perspective. *Acad Med*. 1999;**74**:154-164.
[103] Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;**78**:775-780.
[104] Croskerry P. Cognitive forcing strategies in clinical decisionmaking. *Ann Emerg Med*. 2003;**41**:110-120.
[105] Eva KW, Hatala RM, Leblanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ*. 2007;**41**:1152-1158.
[106] Mayer R. Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*. 2010;**20**:167-171.
[107] Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. *Educ Psychol*. 2003;**38**:43-52.
[108] Boucheix J-M, Lowe RK. An eye-tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*. 2010;**20**:123-135.
[109] Ginns P. Meta-analysis of the modality effect. *Learning and Instruction*. 2005;**15**:313-331.
[110] Raskin RE. Skin and subcutaneous tissues. In: Raskin RE, Meyer DJ, eds. *Canine and feline cytology: a color atlas and interpretation guide*. 2nd edn. Elsevier, St. Louis, MO, 2010; 26-77.
[111] Rosenthal R, Rosnow RL. *Essentials of behavioral research: Methods and data analysis*. 2nd ed. New York McGraw Hill; 1991.
[112] Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ Erlbaum; 1988.
[113] Sherbino J, Dore KL, Wood TJ, et al. The Relationship Between Response Time and Diagnostic Accuracy. *Acad med*. 2012.
[114] Kaufman DR, Yoskowitz N, Patel VL. Clinical reasoning and biomedical knowledge: implications for teaching. In: Higgs J, Jones MA, Loftus S, Christensen N, eds. *Clinical reasoning in the health professions*. 3rd edn. Elsevier, Philadelphia, PA, 2008.
[115] Grant ER, Spivey MJ. Eye movements and problem solving: guiding attention guides thought. *Psychol Sci*. 2003;**14**:462-466.
[116] Thomas LE, Lleras A. Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review*. 2007;**14**:663-668.
[117] Mylopoulos M, Woods NN. Having our cake and eating it too: seeking the best of both worlds in expertise research. *Med Educ*. 2009;**43**:406-413.
[118] Evered A, Walker D, Watt AA, Perham N. To what extent does nonanalytic reasoning contribute to visual learning in cytopathology? *Cancer Cytopathol*. 2013;**121**:329-338.
[119] Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ*. 2010;**44**:94-100.

Appendix 1. Equipment set up.



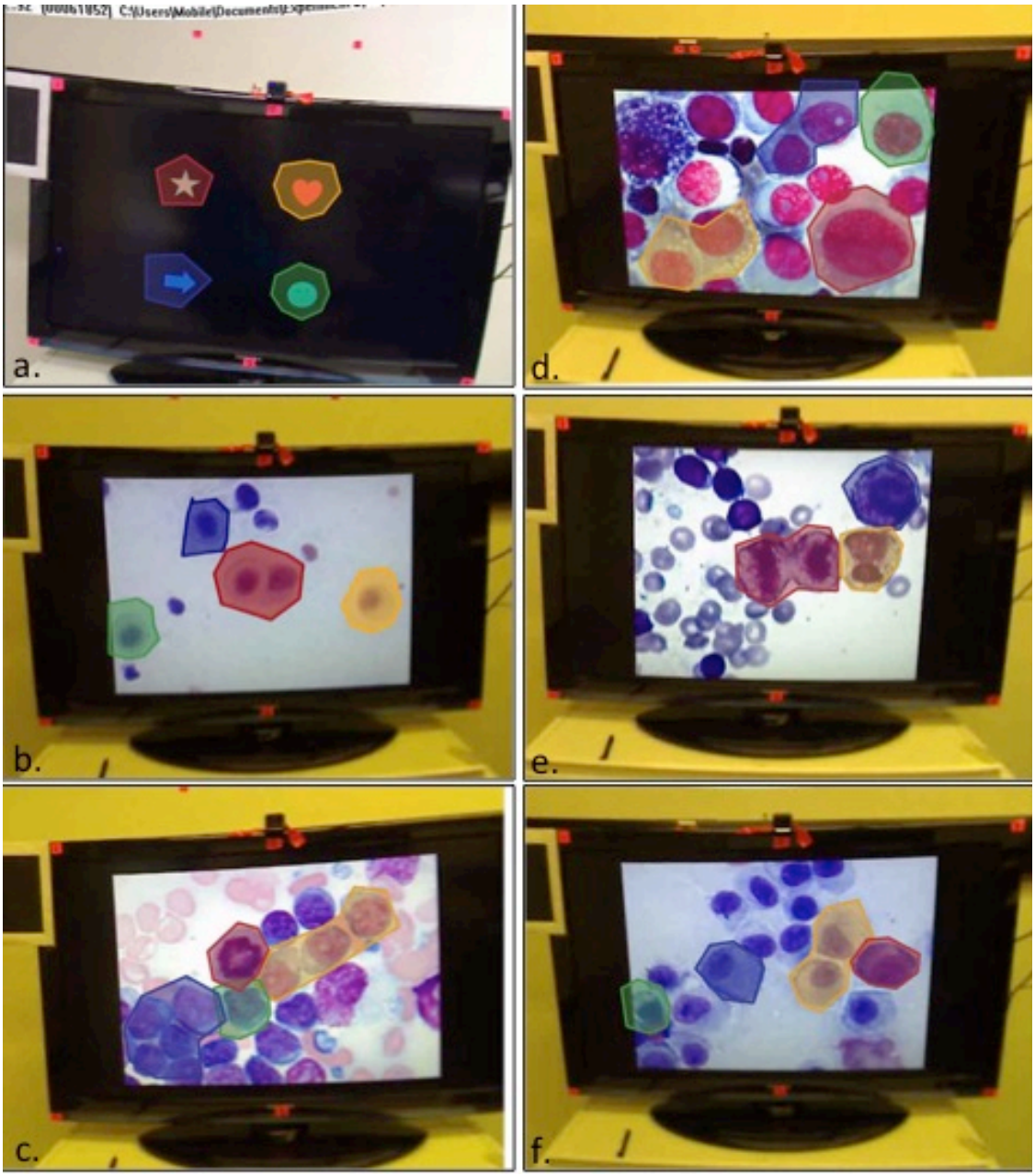
Equipment set up with participant wearing eye-tracking glasses seated approximately 1 meter from the LCD screen projecting a cytology image (Image 1). On the top of the LCD screeen is a web-cam recorder. Behind the participant is the video recorder.

Appendix 2. Eye-tracking calibration set up.



41 inch LCD screen with "calibration slide" displayed and environment reference target in top left corner. Around the periphery of the LCD screen are 8 pink calibration points. The "star", "heart", "arrow" and "smiley face" were used as calibration points on the LCD screen as well as to verify calibration prior to each recording session.

Appendix 3a. Calibration image and images 1-5 environment maps and AOIs.



a. Calibration slide; b. Image 1: cutaneous histiocytoma; c. Image 2: cutaneous lymphoma; d. Image 3: transmissible venereal tumor; e. Image 4: mast cell tumor; f. Image 5: extramedullary plasmacytoma.

Appendix 3b. Calibration image and images 1-6 environment maps and AOIs.

a. Calibration slide; b. Image 6: cutaneous histiocytoma; c. Image 7: extramedullary plasmacytoma; d. Image 8: transmissible venereal tumor; e. Image 9: mast cell tumor; Image 10: cutaneous lymphoma.

Appendix 4. Novice and expert demographic and previous experience questionnaire.

Research Project Title: Educational interventions to improve veterinary pathology visual diagnostic reasoning measured by eye-tracking technology

Sponsor: University of Calgary

Principal Investigator: Dr. Tyrone Donnon

Co-Investigators: Amy Warren, BSc., BVSc. (hons), DACVP, PhD student.

This questionnaire is to collect educational and demographic information for the study. This information is completely confidential and the only people with access to this information are Drs. Amy Warren and Tyrone Donnon.

Name:_____
Date of birth: _____
Gender: Male □ Female □

Education:
Years of post-secondary school education:
Highest degree attained: Diploma □ Bachelors □ Masters □ Doctorate □

Veterinary and pathology training:
Are you a DVM student? Yes □ No □
If yes: 1st year DVM □ 2nd year DVM □ 3rd year DVM □ 4th year DVM □

Are you a qualified DVM? Yes □ No □
If yes, please list degree, institute and year attained:
_____

Have you completed pathology (anatomic or clinical) specialty training?
If yes: Anatomic pathology residency □ Clinical pathology residency □ Pathology research training (MSc/ PhD) □

Are you a Diplomate of the American College of Veterinary Pathologists: Yes □ No □
If yes please specify if anatomic or clinical pathology and year boarded.
_____

Please include any other previous experience with histopathology or cytology (with details of duration and depth of experience):
_____
_____
_____
_____
_____

Appendix 5: Consent form.

CONSENT FORM

Research Project Title: Educational Interventions to Improve Veterinary Pathology Visual Diagnostic Reasoning Measured by Eye-Tracking Technology

Sponsor: University of Calgary

Principal Investigator: Dr. Tyrone Donnon

Co-Investigators: Amy Warren, BSc., BVSc. (hons), DACVP, PhD student.

This consent form is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mention here, or information not included here, please ask. Take the time to read this carefully and to understand any accompanying information. You will receive a copy of this form.

Background

Visual diagnostic reasoning is an essential skill clinicians develop during their veterinary training. Reasoning capability is based on the integration and effective application of thinking and learning skills, to generate knowledge within familiar and unfamiliar clinical experiences. Histopathology and cytology, sub-specialties of veterinary anatomic and clinical pathology respectively, rely on the integration of visual cues from histopathologic/ cytologic specimens to form a diagnosis. Recent studies in human radiology and dermatology have utilized eye (gaze)-tracking technology to assess visual diagnostic reasoning skills. With the development of virtual microscopy utilizing instruments such as the Aperio scope, eye-tracking technology is now available to pathologists to study visual diagnostic reasoning. Using eye-tracking technology, we aim to see if there is a difference between novice (veterinary student) and expert (board-certified specialist) pathologists in eye-movements when reading and diagnosing a virtual microscopic slide.

What is the Purpose of the Study?

The goal of this study is to determine the baseline differences in eye-movement patterns utilizing eye-tracking technology between novice (DVM students) and expert pathologists (ACVP board certified DVMs). Using these baselines we intend to use eye-tracking to assess two educational interventions in teaching visual diagnostic reasoning skills to veterinary students.

What Would I Have to Do?

After you have read and signed the informed consent regarding your voluntary participation in this research study, you will be randomly assigned to a comparison group. Each group will undergo an hour long teaching session on diagnostic cytology. After this time, each student will be asked to perform a 20 minute visual diagnostic exercise recorded using eye-tracking glasses. A short 1 minute calibration will precede the diagnostic exercise. Each session will recorded on DVD and viewed only by Amy Warren and Tyrone Donnon.

What Are the Risks?

There are no greater risks to participation in this study than those ordinarily experienced in daily life. The results will not to be used in the assessment of VETM 421.

Will I Benefit If I Take Part?

Once your results are compiled, the direct benefit to you will be a summary report of your personal results. From an educational perspective, the implications of this work have theoretical and practical applications that may influence teaching and evaluation techniques pertaining to visual diagnostic reasoning training.

Do I Have to Participate?

Your participation in this research study is completely voluntary. You can refuse to participate or withdraw at anytime during the course of the study without retribution.

What Else Does My Participation Involve?

Your participation will also involve signed consent to permit access to data sources from the University of Calgary's DVM program 1) demographics information (name, age, gender, etc.) and 2) Grades in first and second year DVM subjects. Please note that all of the data will be compiled into an aggregated format and published anonymous such that information from any one person will be kept strictly confidential.

Will I Be Paid For Participating, Or Do I Have to Pay For Anything?

You will not be paid for your participation and your involvement will not cost you anything more than the time that is required to complete the educational session and testing.

Will My Records Be Kept Private?

The data will be gathered and processed in such a way as to ensure confidentially and complete anonymity by the Principal Investigator. As such, each participant will be assigned an anonymous Study Identification Number, results will be presented in an aggregated format that will not identified any one person, and all of the research records obtained will be stored together and locked away in an office file cabinet at our facilities in the Health Science Centre. All data will be kept in a secured office inaccessible to others, and all of the collected paper copies and electronic records will be destroyed five years after completion of the study.

If I Suffer A Research-Related Injury, Will I Be Compensated?

In the event that you suffer injury as a result of participating in this research, no compensation will be provided to you by the University of Calgary, the Faculty of Veterinary Medicine or the Researchers. You still have all your legal rights. Nothing said in this consent form alters your right to seek damages.

Signatures

Your signature on this form indicates that you have understood to your satisfaction the information regarding your participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the investigators or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time without jeopardizing your position or state of well-being. If you have further questions concerning matters related to this research, please contact:

Dr. Tyrone Donnon (403) 210-9682

Amy Warren (403) 210-6179

If you have any questions concerning your rights as a possible participant in this research, please contact The Director of the Conjoint Health Research Ethics Board at the Office of Medical Bioethics, 403-220-7990.

Participant's Name	Signature and Date
Investigator/Delegate's Name	Signature and Date
Witness' Name	Signature and Date

The University of Calgary Conjoint Health Research Ethics Board has approved this research study. A signed copy of this consent form has been given to you to keep for your records and reference.

Signatures

Your signature on this form indicates that you have understood to your satisfaction the information regarding your participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the investigators or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time without jeopardizing your position or state of well-being. If you have further questions concerning matters related to this research, please contact:

Dr. Tyrone Donnon (403) 210-9682

Amy Warren (403) 210-6179

If you have any questions concerning your rights as a possible participant in this research, please contact The Director, Office of Medical Bioethics, University of Calgary, 403-220-7990.

Participant's Name	Signature and Date
Investigator/Delegate's Name	Signature and Date
Witness' Name	Signature and Date

The University of Calgary Conjoint Health Research Ethics Board has approved this research study. A signed copy of this consent form has been given to you to keep for your records and reference.

Appendix 6a. Lesson plan for Traditional teaching intervention (Group 1: Didactic teaching)

Title of lesson
Group 1: Control (traditional teaching)

Purpose
Control group- the five round cell tumors are taught in a didactic setting with no active participation of the students. This will be based on a powerpoint.

Pre-test (1 hour)
Eye-tracking of each student in the group using images of the five tumor types to be taught. These data will be used as a baseline to assess improvement of the students' ability to recognize the tumors visually.

Bridge-in (3 minutes)
Image of a mast cell tumor (most recognizable of the 5 tumors) and questions on what the tumor is, why and what visual features they are identifying that allows them to come to that diagnosis.

Input from you (50 minutes)

- Introduction to types of subcutaneous masses
- Cytological algorithm of decisions
- Round cell tumor types
- Visual differentiating features of the 5 tumor types
- Prognosis of tumors
- Medical and surgical interventions for the 5 tumors.

Closure (3 minutes)

Re-iterate the diagnostic features of the 5 round cell tumors

Guided practice (application of knowledge: classroom activities for students, problem to solve, etc.)

None

Check for understanding (1 hour)

Post-test eye-tracking of the five tumor types.

Appendix 6b. Basic visual reasoning teaching intervention (Group 2: Active learning, single image)

Title of lesson
Group 2: Active participation with a single static image.

Purpose ("why" of the lesson, where and how does it fit into the course/curriculum)
To assess if allowing students active participation by providing a single example of a tumor type in a static image improves visual diagnostic reasoning more than no active participation (control)

Pre-test (1 hour)
Eye-tracking of each student in the group using images of the five tumor types to be taught. These data will be used as a baseline to assess improvement of the students' ability to recognize the tumors visually.

Bridge-in (3 minutes)
Image of a mast cell tumor (most recognizable of the 5 tumors) and questions on what the tumor is, why and what visual features they are identifying that allows them to come to that diagnosis.

Input from you (20 minutes)

- Introduction to types of subcutaneous masses
- Cytological algorithm of decisions
- Round cell tumor types
- Visual differentiating features of the 5 tumor types

Closure (10 minutes)

The 5 static images (one of each tumor) will be reviewed by the instructor to make sure the students were able to differentiate the tumors and identify the visual diagnostic features.

Guided practice (20 minutes)

Students will be provided with a single static image of each tumor type. Using the key features described they will be instructed to compare and contrast visual features of each tumor to come up with a diagnosis.

Check for understanding (1 hour)

Post-test eye-tracking of the five tumor types.

Appendix 6c. Extended visual reasoning teaching intervention (Group 3: Active learning, image repetition)

<table>
<tr><td colspan="2">Title of lesson
Group 3: Active participation with image repetition</td></tr>
<tr><td colspan="2">Purpose (“why” of the lesson, where and how does it fit into the course/curriculum)
To assess if practice with 5 static image examples of each of the five tumor types improves visual diagnostic reasoning as compared to the control (traditional teaching) and single static image.</td></tr>
<tr><td colspan="2">Pre-test (1 hour)
Eye-tracking of each student in the group using images of the five tumor types to be taught. These data will be used as a baseline to assess improvement of the students’ ability to recognize the tumors visually.</td></tr>
<tr><td colspan="2">Bridge-in (3 minutes)
Image of a mast cell tumor (most recognizable of the 5 tumors) and questions on what the tumor is, why and what visual features they are identifying that allows them to come to that diagnosis.</td></tr>
<tr><td colspan="2">Input from you (10 minutes)
• Round cell tumor types
• Visual differentiating features of the 5 tumor types</td></tr>
<tr><td>Closure (10 minutes)

The 25 static images (five of each tumor) will be reviewed by the instructor to make sure the students were able to differentiate the tumors and identify the visual diagnostic features.</td><td>Guided practice (35 minutes)

Students will be provided with a five different static images of each tumor type (25 in total). The images will be presented in a mixed (rather than blocked) practice model. Using the key features described they will be instructed to compare and contrast visual features of each tumor to come up with a diagnosis.</td></tr>
<tr><td colspan="2">Check for understanding (1 hour)
Post-test eye-tracking of the five tumor types.</td></tr>
<tr><td colspan="2"></td></tr>
</table>

Appendix 7. Gaze-map calculated statistics for eye-tracking data for expert 4, image 1.

	# Subject Video Information:			
Total Length of Video (hours:minutes:seconds / frames)	0:00:16.683 / 500			
Statistics computations started (hours:minutes:seconds / frames)	0:00:00.66 / 3			
Statistics computations ended (hours:minutes:seconds / frames)	0:00:16.716 / 502			
Percent Confidence in Subject Camera File	99.80%			
# AOI Summary Information:				
Number of AOIs	4			
First AOI fixated	AOI 2 Bottom LH LSA			
Percent of Time Spent in Any AOI	10.40%			
Percent of Time Spent Outside Any AOI	89.60%			
Percent of Time Spent Outside Image View	2.00%			
Region	AOI 0 Top RH LSA	AOI 3 Bottom RH LSA	AOI 2 Bottom LH LSA	AOI 1 Top central LSA
Number of Gaze Points	2	0	32	18
Time in Region (seconds)	0.07	0	1.07	0.6
Number of Dwells	1	0	4	5
Average Dwell Duration (seconds)	0.07	0	0.27	0.12
Time to First Gaze (seconds)	8.07	-0.03	1.87	8.91
Percent Gaze In Region To Total Gaze	0.4	0	6.4	3.6
Percent Gaze in Region to in Any AOI	3.85	0	61.54	34.62
Number of Fixations	0	0	2	2
Average Fixation Duration (seconds)	0	0	0.38	0.15
Total Fixation Duration (seconds)	0	0	0.77	0.3
Time to First Fixation (seconds)	-0.03	-0.03	2.04	14.71
Percent Fixation Time to Dwell Time	0	0	71.88	50
Average Pupil Diameter in Region (pixels)	73.49	0	75.64	72.61
Std. Dev. Pupil Diameter in Region (pixels)	0.06	0	1.37	0.84

Appendix 8. Think-aloud raw data for image 7 (extramedullary plasmacytoma)

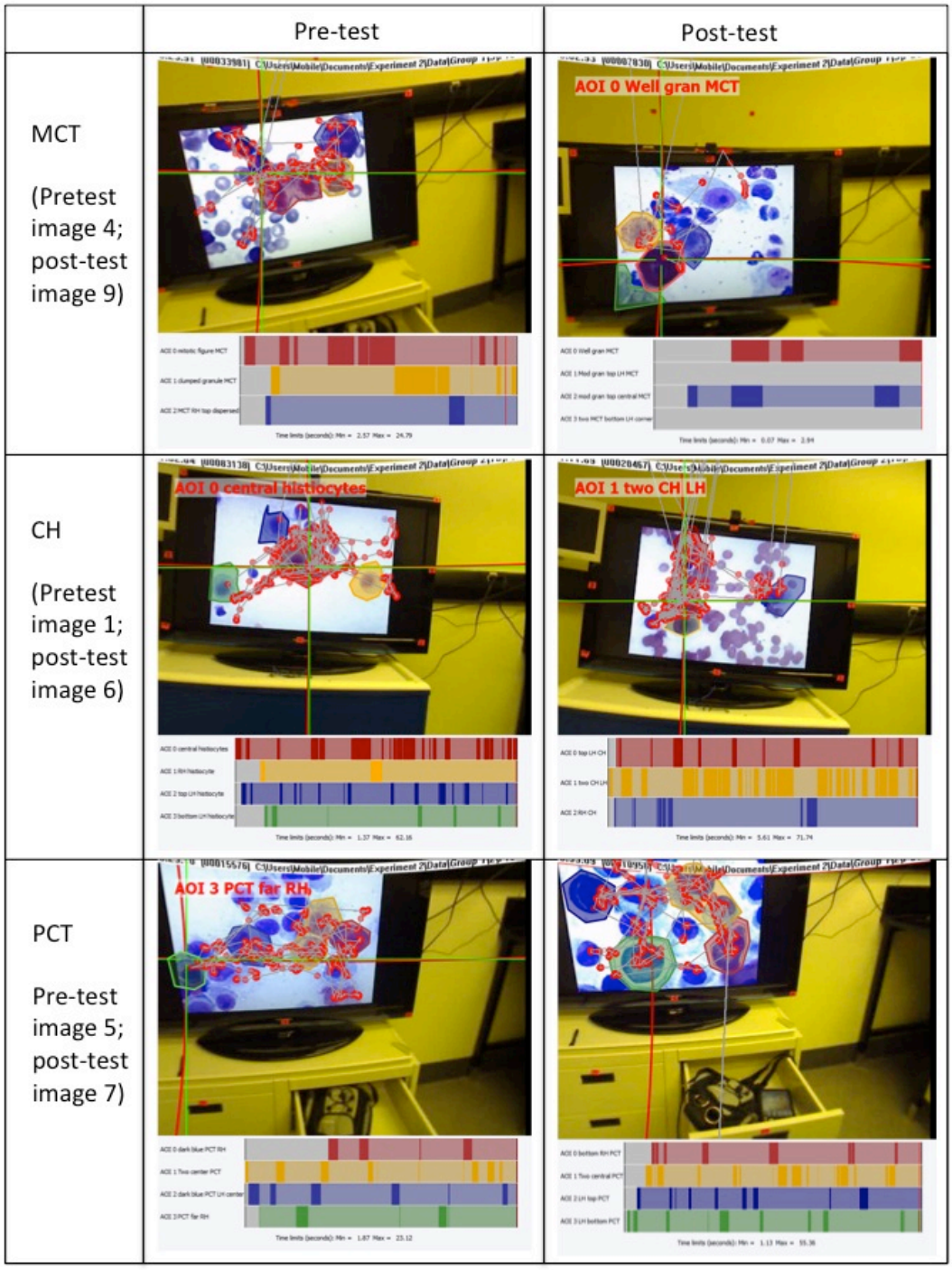
Expert identified features	ID	Perinuclear clear zone**	Deep blue cytoplasm**	Eccentric nuclei**	Distinct round cells*	Clumped chromatin	Correct?
Novice 1	20		x	x			0
	23						0
	2						0
	19						0
	30						0
	13						0
Novice 2	3			x			0
	10						0
	17						0
	15						0
	27		x				0
Total Novice		0	2 (18.2)	2 (18.2%)	0	0	0
Expert 1	4	x	x				1
	12	x	x	x			1
	25	x	x	x			1
	24	x	x	x	x		1
Expert 2	8	x	x	x			1
	5	x	x		x	x	1
Total Expert		6 (100%)	6 (100%)	4 (66.6%)	2 (33.3%)	1 (16.6%)	6

*Distinct round cells excluded as a feature common to round cell tumor group
**Perinuclear clear zone, deep blue cytoplasm and eccentric nuclei identified as "key" diagnostic features based on frequency identified by experts (>50.0%)

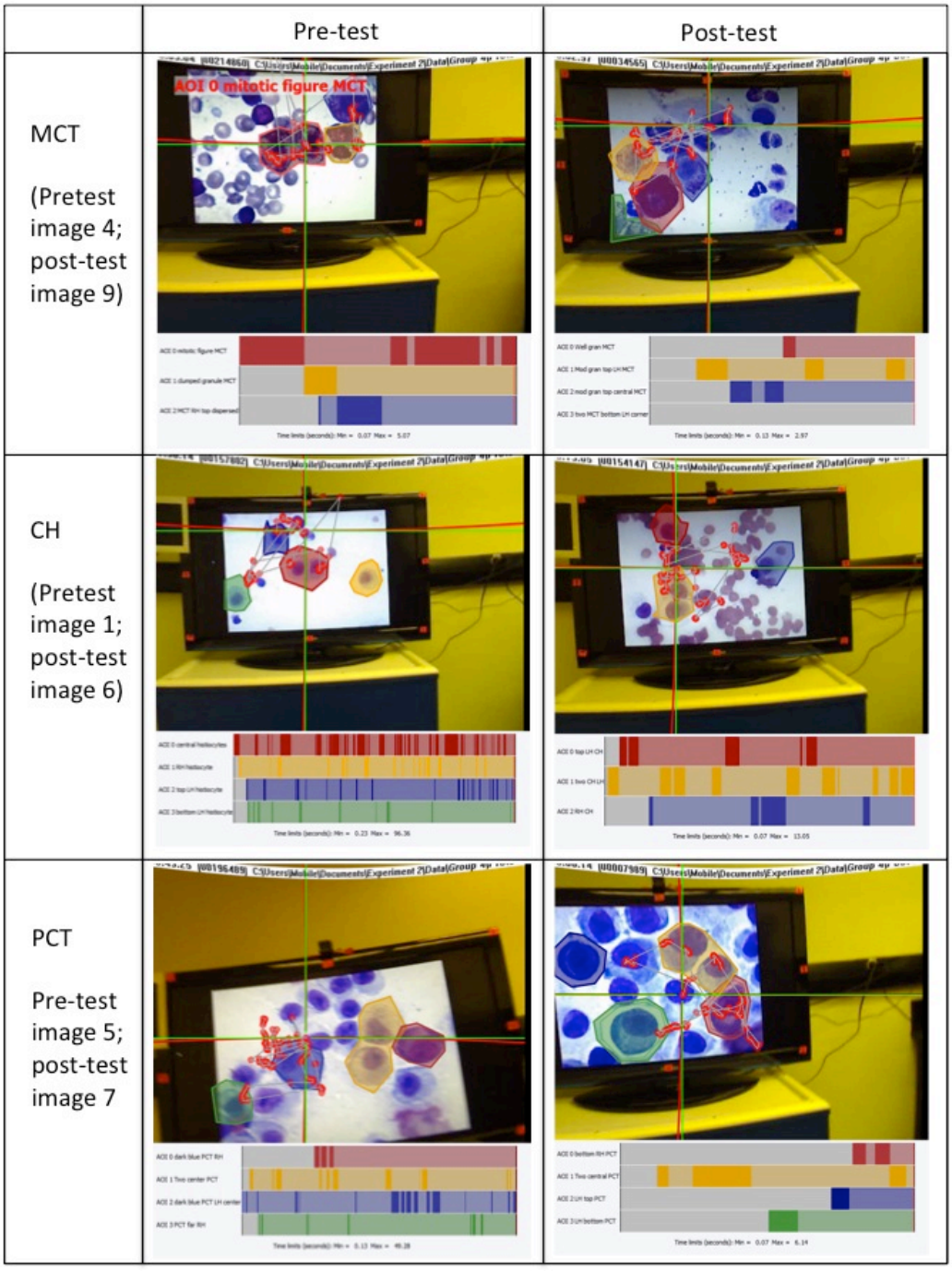
Other incorrect or irrelevant features identified by novices and not identified by experts

- 2- cluster of cells
- 2- no RBC
- 13- cells all the same
- 23- not uniform
- 23- very blue – too much stain
- 30- stained background
- 30- no RBC
- 3- prominent nucleoli
- 10- cells larger than RBC
- 10- large nucleoli
- 15- uniform cells
- 17- not uniform

Appendix 9a. Pre- and post-test eye-movement patterns for teaching intervention group 1 for the mast cell tumor (MCT) (student 3), cutaneous histiocytoma (CH) (student 17), and extramedullary plasmacytoma (PCT) (student 2).



Appendix 9b. Pre- and post-test eye-movement patterns for teaching intervention group 2 for the mast cell tumor (MCT) (student 29), cutaneous histiocytoma (CH) (student 6), and extramedullary plasmacytoma (PCT) (student 8).



Appendix 9c. Pre- and post-test eye-movement patterns for teaching intervention group 3 for the mast cell tumor (MCT) (student 25), cutaneous histiocytoma (CH) (student 18), and extramedullary plasmacytoma (PCT) (student 7).