2013-10-18

Case Specificity and Intervening Variables Affecting Diagnostic Accuracy within Two Similar Clinical Presentations

Munshi, Fadi

UNIVERSITY OF CALGARY

Case Specificity and Intervening Variables Affecting Diagnostic Accuracy within Two Similar Clinical Presentations

by

Fadi Mohammad Munshi

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 SCIENCE

CALGARY, ALBERTA

OCTOBER, 2013

Abstract

This study investigated the generalizability of diagnostic performance and problem solving strategy within and across cases from two similar clinical presentations (CP). Many variables have been described in medical education literature that influences the generalizability of diagnostic performance. This doctoral research further examined the effect of intervening variables in diagnostic performance. The variables studied were: disease difficulty, case typicality, CP, and problem solving strategy.

Participants of the study were sixth year medical students enrolled in four universities in Saudi Arabia. Two types of data were analyzed, diagnostic performance scores obtained from written clinical vignettes and problem solving strategies abstracted from written think aloud exercises.

Classical Test Theory (CTT) and Item Response Theory (IRT) were two measurement frameworks used to analyze the data and triangulate study findings. Variable reduction strategies were used to examine the construct of performance scores in comparison to the CP schemes. Cross-tabulation and Chi-square were used to investigate the consistency in using problem solving strategies within and across cases from two similar CPs. Also, Multi-Facet Rasch Modelling (MFRM) was employed to determine degree of impact on diagnostic performance by the independent variables.

This study answered three research questions related to diagnostic generalizability and intervening variables that contribute to diagnostic performance. Most of the diseases included in this study showed high correlation coefficients and high discriminating measures. The estimated measures from two different measurement theories supported predictability of performance from one case to another.

This research provides evidence that the diagnostic strategy utilized can alter diagnostic performance. Further support was found for schema-based instruction in medical schools to enhance student diagnostic performance.

The research model illuminates some factors that influence diagnostic performance. Disease difficulty and type of CP were two variables found to alter diagnostic performance. Further research is required to identify other variables.

Finally, these study findings set the base for future research. Directions for further investigation include: examining performance and problem solving strategy generalizability within and across dissimilar CPs; studying the effect of case typicality on diagnostic performance using MFRM; and including a broader sample that would include participants from medical schools that have different curricular models.

Preface

The purpose of this dissertation was to investigate the generalizability and intervening variables that affect diagnostic performance in the context of CPs. This doctoral research should be of interest to medical schools, medical councils and post-graduate training programs. The work included in the dissertation is original, unpublished, and independent work by F. Munshi.

Acknowledgements

I would like to give my deepest gratitude to Dr. Harasym, my PhD supervisor, for his guidance, patience, and encouragement during the PhD journey. He has been a great mentor for me not only in graduate studies but also in life.

My sincere thanks go to Dr. Beran and Dr. Baig, for their directions on my academic works throughout my doctoral study and for all the support during the hard times. I sincerely appreciate Dr. McIlwrick, Dr. Kassam, and Dr. Papa for their time and comments on my dissertation.

Special thanks go to my wife, Aisha, for her support, encouragement, and love; to my sweet son Hamzah and cute daughter, Saliha, for being healthy and happy children and having brought me the most amazing experience in my life; to my father, siblings, and in-laws for always being so supportive to me in countless ways. Gratitude is extended to the clerks at the four participating universities, King Abdulaziz, Umm Al-Qura, Taibah, and Qassim.

Dedication

This dissertation is lovingly dedicated to the memory of my mother, Saliha Atiah. Her support, encouragement, and constant love have sustained me throughout my life. Mom, you will always be the never ending source of comfort, happiness, and being in my heart.

Table of Contents

List of Tables

List of Figures

Definition of Terms

Term & Acronyms	Definition
Abstract Theory	Categorization theory that suggests physicians form generalizations about how disease classes present. These generalizations take the form of conditional probability estimates.
Analysis of Variance (ANOVA)	Statistical models used to analyze the differences between group means and their associated procedures.
Case Specificity	Diagnostic performance on one clinical case does not predict performance on another case.
Classical Test Theory (CTT)	A body of related psychometric theory that predicts outcomes of psychological testing. CTT assumes that each person has a *true score* that would be obtained if there were no errors in measurement.
Clinical Presentation (CP)	The common and important ways in which a person, group of people, community, or population present to a physician.
Clinical Vignette	A concise, written case history of a fictitious patient based on a realistic clinical situation that is accompanied by one or more questions that explore what a physician would do if presented with the actual patient.
Declarative Knowledge	Factual information stored in memory.
Diagnostic Competency	A set of defined behaviours that provide a structured guide enabling the identification, evaluation and development of the behaviours in individual physicians
Diagnostic Performance	The ability to recognize a patient`s disease or condition by signs and symptoms.
Differential Diagnosis	A systematic diagnostic method used to identify the presence of an entity where multiple alternatives are possible.
Exemplar Theory	Categorization theory that assumes physicians store knowledge of signs and symptoms experienced with each patient encountered.

Expertise	Knowledge, skills, and characteristics that are acquired in a certain domain through training and experience.
Extended matching questions (EMQ)	A written examination format similar multiple choice questions but with one key difference, that they test knowledge in a far more applied, in depth, sense.
Factor Analysis (FA)	A statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.
Generalizability	The ability to predict diagnostic performance from one clinical case to another.
Guessing	To predict without sufficient information.
Hypothetical Deductive (HD) Reasoning	An approach that begins with the diagnosis then works back to signs and symptoms.
Item Response Theory (IRT)	A paradigm for the design, analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other variables. Unlike simpler alternatives for creating scales evaluating questionnaire responses it does not assume that each item is equally difficult.
Multiple Choice Question (MCQ)	A form of assessment in which respondents are asked to select the best possible answer out of the choices from a list.
Multi-Facet Rasch Modelling (MFRM)	A psychometric model for analyzing data, as a function of the trade-off between (a) the respondent's abilities and (b) the item difficulty.
Pattern Recognition	A reasoning strategy in which an expert further establishes key features of diseases in a scheme that enables recognition of diseases.

Principal Component Analysis (PCA)	A mathematical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
Procedural Knowledge	Knowledge of how to perform a task or procedure.
Scheme	Highly organized knowledge in which concepts are arranged from general signs/symptoms to specific diseases.
Scheme Inductive (SI) Reasoning	A heuristic approach that utilizes a highly organized knowledge structure and forward reasoning to reach a diagnosis.
Simulation	Approximation to reality that requires trainees to react to problems or conditions as they would under genuine circumstances.
Standardized Patient (SP)	An individual who is trained to act as a real patient in order to simulate a set of symptoms or problems.
Typicality Gradient	A continuum of patient presentations within a certain disease ranging from prototypical (usual signs and symptoms) to atypical (not usual signs and symptoms).
Written Think Aloud (WTA)	A data gathering method that involves participants writing their thinking process as they are or after performing a set of specified tasks.

Epigraph

“To Err is Human but to Be Knowledgeable is Humane”

Peter H. Harasym, Ph.D.

Chapter One: Introduction

1.1 Overview

Errors in healthcare are a leading cause of death and injury. An error can be defined as "the failure of a planned action to be completed as intended (i.e., error of execution) or the use of a wrong plan to achieve an aim (i.e., error of planning)" (Kohn, Corrigan, & Donaldson, 2000). Leape (1994) characterized the kinds of errors that resulted in medical injury as diagnostic, treatment, preventive, or other errors such as communication failure. Diagnostic errors are categorized into errors or delay in diagnosis, failure to employ indicated tests, use of outmoded tests of therapy, and failure to act on results of monitoring or testing. The focus of this study is to investigate factors that could contribute to diagnostic errors (i.e., factors that enhance diagnostic performance). This study also investigated the extent that performance generalizes/predicts diagnostic performance within cases from one clinical presentation (CP) to cases from another similar CP.

The Accreditation Council for Graduate Medical Education guidelines requires any medical program to integrate six competencies into the curriculum. Physicians must demonstrate medical knowledge, patient care, communication, practice-based learning, professionalism, and system-based practice competencies in their attainment of competency in diagnosis. Evidence of diagnostic success is demonstrated by the ability to correctly diagnose a given disease despite the many different combinations of signs and symptoms with which the disease may manifest (Papa & Elieson, 1993).

1.1.1 Categorization Theories

Competency in differential diagnosis is the cornerstone of clinical medicine. Medical cognition research suggests that diagnostic performance is better understood by categorization

theories. The exemplar and the abstraction theories are examples of categorization theories (Papa, Stone, & Aldrich, 1996).

Exemplar theory assumes that physicians store knowledge of signs and symptoms experienced with each patient encountered. When confronted with a new case, they try to link the new signs and symptoms to those that best match previous encountered patients. In abstraction theory, it is suggested that physicians form generalizations about how disease classes present. These generalizations may take the form of conditional probability estimates; for example, the percentage of patients with a myocardial infarction that present with chest discomfort. These conditional probabilities are transformed into weights to distinguish between various diseases that present with a given CP. The disease differential with the highest accumulated weight is the diagnosis for the new patient encounter. In exemplar theories, the main diagnostic strategy is believed to be hypothetical deductive (HD) reasoning (matching of diseases to presenting signs and symptoms). In abstraction theories each sign/symptom has an associated weight for each of the various diseases associated with a given presentation. The different weightings that a single sign or symptom carries for each of the various disease differential under consideration enables reasoning across multiple diseases as might a schema. In essence, these various disease weightings for a given sign/symptom represent a given heuristic. Thus, each sign/symptom and its respective disease weightings represent 'one heuristic' and thereby enables forward reasoning one sign/symptom at a time. Contrary to the schema theory description below, these weightings are held in working memory and the cumulative weights at any given point in the work up (elicitation of signs/symptoms) enable the intelligent derivation of the optimal signs and symptoms that should be gathered from that point forward.

A third theory that can explain diagnostic performance is the scheme theory. Given a CP, the scheme theory involves a heuristic approach that utilizes a highly organized knowledge structure and forward reasoning (scheme induction) to reach a diagnosis (Bordage, 1994; Schmidt, Norman, & Boshuizen, 1990). Inductive theory suggests that there is no matching or weighting of signs and symptoms until the final diagnosis is made or confirmed. In scheme theory, the movement is forward through an organized knowledge structure, which may terminate with HD reasoning at the very end of diagnostic reasoning.

1.1.2 Enhancing Diagnostic Performance within Medical Curricula

Many medical schools' educational experiences are organized in accordance to the disciplinary, body systems, and problem-based clinical models. The vast majority of the medical schools still use the disciplinary model that originated in Europe and was implemented in Northern America in 1871 (Papa & Harasym, 1999). All three curricular models teach students hypothetical-deductive reasoning (a "backward" cognitive reasoning approach). Experts are more inclined to use scheme-inductive reasoning and pattern recognition that is based on a highly organized body of basic and clinical knowledge that permits "forward" clinical reasoning (McLaughlin, Coderre, Mortis, & Mandin, 2007)

Theoretically, it can be assumed that generalizability is influenced by the degree of advanced knowledge organization. That is, experts with highly organized knowledge structures will use either SI reasoning/pattern recognition to arrive at a diagnosis (Bordage, 1994; Harasym, Tsai, & Hemmati, 2008). These experts are more likely to obtain the correct diagnosis within and across levels of a CP than physicians whose knowledge is not highly organized. Highly organized knowledge is hierarchical (concepts are arranged from general signs/symptoms to

specific diseases), and has at least 3-5 levels of branching. The structure has logical branch points that allow the physician to move in a forward manner from presenting complaint to diagnosis (McLaughlin, Coderre, Mortis, & Mandin, 2007). An organized medical knowledge structure is often called a scheme (Mandin, Harasym, Eagle, & Watanabe, 1995).

1.2 Problem Statement

Diagnostic performance of a physician is determined by the ability to diagnose a disease regardless of how it presents in terms of various combinations of signs and symptoms. A physician may see cases that range from a prototypical to an atypical presentation. Case specificity suggests success in diagnosing a case is specific to that case and cannot be generalized to other cases (Norman, Bordage, Page, & Keane, 2006). Therefore, an expert in chest pain may not accurately diagnose a patient with back pain. For each chief complaint, there is a unique body of declarative and procedural knowledge. This implies that the assessment of diagnostic performance must be performed using a number of different case presentations for a specific disease, as success in diagnosing a clinical case is content/knowledge dependent and process dependent (Harasym et al., 2008; Papa, Oglesby, Aldrich, Schaller, & Cipher, 2007).

However, it is unknown whether the generalization in diagnostic performance will occur within a CP – that is, the predictability of performance above levels, across levels, and between CPs. Therefore, there is a need to investigate if generalization exists. With the content and case specificity in mind, we chose to study two similar CPs, chest discomfort and dyspnea.

1.3 Significance of the Research

Due to the content and case specificity phenomena, it is assumed that generalization will not occur across CPs. Should generalization of diagnostic accuracy occur within a CP, number of

cases required to reliably assess diagnostic performance can be significantly reduced if the degree of generalizability is high. Conversely, if there is little generalization of diagnostic performance within a CP, then the number of cases required to reliably assess diagnostic performance would be increase accordingly.

1.4 Conceptual Structure of Intervening Variables

This study investigated the extent to which diagnostic performance generalizes within and across two similar CPs and the variables that may enhance diagnostic accuracy. This study will benefit licensing agencies and medical schools by clarifying the factors and the number of items needed to assess diagnostic performance within two similar CPs.

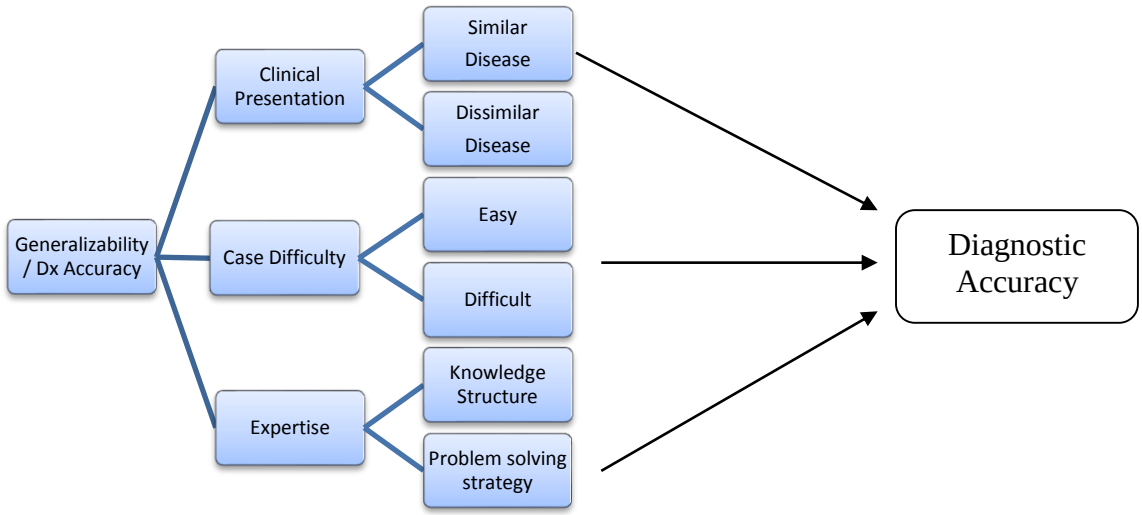


Figure 1. A model illustrating the hypothesized intervening variables that are anticipated to influence the generalizability of diagnostic performance within and across CPs.

Intervening variables that may affect the degree of generalizability include: the disease assessed within a CP, examinee ability, disease difficulty and examinee's diagnostic problem solving strategy. Figure 1, illustrates the conceptual structure of the intervening variables that could affect generalizability of diagnostic accuracy.

1.5 Study Purpose

The purpose of this doctoral research was: (1) to investigate the degree to which diagnostic performances will generalize within a similar cluster of diseases (i.e., CP) and across a dissimilar cluster and (2) to identify the factors that are anticipated to influence diagnostic performance generalization.

The thesis is divided into five chapters. Chapter one provides background information and states the purpose of this study. Chapter two provides a review of literature relevant to the current study, ending with the research questions that guide this scholarly work. Chapter three outlines the research methods utilized to answer the research questions mentioned in chapter two. Chapter four presents the results of the study. Finally, chapter five restates research results in a discussion form for the purpose of evaluating the research questions. The chapter ends with recommendations for future research.

Chapter Two: Review of Literature

2.1 Introduction

Many variables have been described to influence the generalizability of diagnostic performance. This review will focus on three main intervening variables, which emerge repeatedly throughout the literature. The three main variables are: CPs, disease typicality, and candidate expertise. Although the literature presents these variables in a variety of contexts, this review of literature will primarily focus on their application in medical education.

2.2 Clinical Presentations

According to cognitive science, humans mentally construct their own learning and build new schemata, or structures that categorize comprehension, as they experience new information and process the stimuli that are received by the brain. They also use prior knowledge, social experiences, and emotions as they subconsciously select which sensory information to accept or adapt, and which to ignore (Fox, Phung, & Waycott, 2006).

An innovative model, clinical presentation curriculum (CPC), is based on the assumption that the manner or mode in which the human body reacts to an infinite number of insults is finite and stable over time (Mandin, Harasym, Eagle, & Watanabe, 1995). The manner patients present to physicians are termed CPs, defined as the common and important ways in which a person, a group of people, a community, or population present to a physician. In CPC, teaching is organized around 120-130 ways a patient presents to a physician (e.g., chief complaint). These CPs take the form of signs (e.g., hypertension), chief complaint (e.g., chest discomfort), and/or laboratory abnormalities (e.g., elevated serum lipids). Each CP has a unique scheme that

integrates both basic and clinical sciences teaching, which aids learning and clinical problem solving (Mandin, Harasym, & Woloschuk, 2000).

2.3 Case Typicality

When patients with the same disease present a physician to seek medical attention, they may present with different combinations of signs and symptoms. Papa et al. (1996) referred to the varying combinations of signs and symptoms as case typicality gradient. Typicality gradient is a continuum in which presentations of the same problem range from prototypical (easy) to atypical (difficult). In this study, Papa and his colleagues concluded that with a typical case presentation, participants were more likely to diagnose the case properly (Papa et al., 1996). The process of defining the typicality level of a clinical vignette is complicated, however Papa and colleagues have demonstrated through the use of abstraction theory and artificial intelligence research to develop eight levels of typicality for diseases of chest pain (Papa et al., 1996; Papa, Harasym, Schumacker, Aldrich, & Stone, 1998).

Given the wide variety of presenting signs and symptoms possible within a given case, and the persistence of the case specificity phenomena, it appears that assessment of diagnostic performance requires a set number and varying typicality of cases within disease classes (Case, Swanson, & Stillman, 1988; Elstein & Schwarz, 2002).

Starting from medical school and throughout their careers, health care professions are exposed to a wide variety of assessments. These target the evaluation of knowledge, clinical skills, and/or attitude.

2.3.1 Clinical Simulations

Simulations are used in the health professions to assess several aspects of clinical competence. They have been integrated into many high-stakes exams including and not limited to the United States Medical Licensing Exam (USMLE), Medical Council of Canada, and Royal College of Physicians and Surgeons of Canada (Dillon, Boulet, Hawkins, & Swanson, 2004). Simulations are defined as "approximations to reality that require trainees to react to problems or conditions as they would under genuine circumstances" (Tekian, McGuire, & McGaghie, 1999). A simulation, whether it involves standardized patients (SPs), computerized case management scenarios, mannequins, clinical vignettes, or a combination of these methods holds great promise for both low-stakes tests in medical schools and for high-stakes licensing and certification assessment (Dillon et al., 2004; Hardie, 2008; Peckler, Schocken, & Paula, 2009). The role of written low fidelity simulations is widely used for assessment in medical education.

2.3.2 Written Assessment Methods

Written assessment methods can be subcategorized by their stimulus format and response format. The stimulus format is what the question asks. The response format is the method the answer is recorded.

I. Stimulus Format

The stimulus format is how the question is asked or presented. The format of stimulus is important in determining the type of competency being evaluated (Schuwirth & van der Vleuten, 2004). If the objective of the test is to examine knowledge recall, then context free questions may be suitable. When higher cognitive thinking, problem solving, and application of knowledge is expected, then context-rich questions are most valuable.

Stimulus format can be categorized into non-, short and long vignette. The variability in these three categories is the level of interpretation, analysis, and synthesis of findings needed to answer the question accurately (Case, Swanson, & Becker, 1996).

a. Non-Vignette Item

The item provides minimal information that is mostly interpreted with no specific patient description. For example, the question would be: What is the first line of treatment for a severe asthma attack? The non-vignette items tend to measure recall/memorization of medical information.

b. Short Vignette Item

A patient's signs and symptoms are described by including only relevant information for interpretation. This format requires a step more in the reasoning process than the non-vignette. In a non-vignette item an example of a question would be: "What is the first line of treatment for a severe asthma attack?" A short vignette item would include description of a patient with a severe asthma attack instead of stating the diagnosis in the item stem.

c. Long Vignette Item

The long vignette will describe a clinical scenario in a more detailed, realistic manner. It will include both relevant and irrelevant information. Relevant information pertinent to making a diagnosis and/or selecting a treatment may include age, gender, medical history, physical examination, and laboratory findings. Irrelevant information may include information unrelated to diagnosis and treatment. Case et al. (1996) investigated the psychometric characteristics of 36 items in non-vignette, short vignette and long vignette formats. The study included 2,643 examinees taking step 2 of the USMLE. The investigators expected the long vignette format to be more difficult and discriminating than the short and non-vignette. They found that the long

vignette format was more difficult when compared to the other two formats. Also, long and short format were more discriminating than non-vignette format. The authors recommended the use of long vignettes, as they are more realistic and require higher cognitive skills to answer. Long vignettes also improve the content and construct validity of exam scores (Case et al., 1996).

II. Response Format

The response format entails how the answer is recorded. In written assessments, the response format is usually either multiple choice questions (MCQ) or open-ended questions. Open-ended questions should only be used when a MCQ cannot be utilized to answer the question. Open-ended questions require more resources (i.e., hand marking) and tend to be less reliable (Schuwirth & van der Vleuten, 2004).

MCQs can be divided into two categories of items: those that require the examinee to indicate the appropriate responses (true/false) and those that require the examinee to select a single response (e.g., one best answer).

a. True/False Family

The true/false family is represented by several specific formats to which letters are assigned to label the item formats. The labeling letters do not represent any specific meaning. The true/false family include C, K, and X formats. C-type items include a set of numbered words or phrases. The examinee is required to decide if A is true, B is true, all are true, or neither is true for each given phrase/statement. This format was designed to compare two diseases, laboratory findings, treatments, etc. The shortcoming of this format is that examinees are challenged to determine the extent to which a certain phrase/statement is true or false. This format can lead the examinee to guess the intention of the examiner rather than test his/her knowledge/ability (Case & Swanson, 2001). This effect reduces the validity of the responses.

K-type items include a stem followed by four alternatives, of which one or more are correct. This format is more difficult and less discriminating in comparison to other formats. It also has the disadvantage of cueing the examinee to select the correct answer (Harasym, Norris, & Lorscheider, 1980). Furthermore, it has lower relative reliability per unit of test time. X-type items are similar to the K-type format without the coded key. Fowel et al. (2000) described the multiple true/false questions as being widely used in the United Kingdom (Fowell, Maudsley, Maguire, Leinster, & Bligh, 2000). True/false questions are criticized for two major shortcomings: guessing and the cueing effect. They are no longer used in the medical licensing and certifying authorities in the United States, as they are difficult to write, may seem ambiguous, and often inappropriately designed to test recall of isolated facts (Case & Swanson, 2001).

b. One Best Answer Family

The one best answer format includes various forms of the responses in which one or more correct responses are selected from a list of options. The response options can vary from 3-5 possibilities up to 20-30 in the case of extended matching questions (EMQs). The one best answer family is also symbolized by capital letters to label specific formats. The one best answer family includes the A, B, R type items. In the A-type format, a single or a set of items are preceded with 4 or more options that challenge the examinee to choose the one best answer. In the B-type, 4-5 matching option items are presented in sets of 2-5 items.

In R-type EMQs, several items are set within a single option list. Each item in a set has a stimulus that can be in the form of a clinical vignette. The examinee is required to select the most likely diagnosis from the list of diseases that may extend to 2-20 options. Case and Swanson (1993) described EMQs as more difficult, more reliable, and more discriminating. Increasing the

number of options in the response possibilities effectively eliminates the cueing effect up to the limit of 20 options. Adding more options will not significantly adjust for the cueing effect (Schuwirth, Van Der Vleuten, Stoffers, & Peperkamp, 1996; Veloski, Rabinowitz, & Robeson, 1993).

2.3.3 Clinical Vignettes

The literature on clinical vignettes in medical education can be grouped into two main categories. The first category of studies investigates and applies clinical vignettes as an assessment tool. The second group of studies utilize clinical vignettes as an instructional instrument.

I. Assessment using Clinical Vignettes

Assessing physicians during daily work is difficult, as they are exposed to a variety of cases that differ in urgency, complexity, and comorbidity. Peabody et al. (2004) investigated the use of clinical vignettes in measuring the quality of clinical practice. This multisite study compared the quality of physician performance using three methods; physicians' performance while caring for unannounced SPs, what physicians wrote in the medical record for the SP visits, and the performance and decisions of physicians while using computer based clinical vignettes of similar patients. Using identical quality measures, differences in physician performance were significantly different but small. Vignettes also measured accurately unnecessary/improper care. The findings were consistent across all diseases, independent of complexity and physician training level and suggest that clinical vignettes are a valid tool for measuring the quality of clinical practice.

Tsai et al. (2009) used clinical vignettes of common ethical dilemmas paired with think aloud interviews to assess expertise in solving ethical problems. The strength of clinical vignettes is that they provide an identical stimulus to all physicians, so that performance can be compared

through various sites and the results are totally attributed to the physician (Norcini, 2004). Clinical vignettes do not identify the effect of other attributes related to health care, such as: teamwork, system-based practice, and the influence of health care systems. The effectiveness of clinical vignettes is highly related to the motivation and ability of participating physicians. Good performance in clinical vignettes does not guarantee high quality patient care while in practice, but poor performance should warrant attention to the physician`s quality of clinical care (Norcini, 2004).

The use of clinical vignettes in evaluating the quality of clinical practice makes it possible to control for case mix, attribute the findings completely to the physician, present low prevalence problems, as well as save on cost and time. A notable disadvantage is the little value clinical vignettes carry in assessing skills like communication and physical examination (Veloski, Tai, Evans, & Nash, 2005; Peabody, Luck, Glassman, Dresselhaus, & Lee, 2000).

2.4 Expertise

Expertise is defined as knowledge, skills, and characteristics that are acquired in a certain domain through training and experience (Ericsson, Krampe, & Tesch-R, 1993). An expert can be a highly experienced professional who gains competencies through instruction and practice. This may include any field such as medicine, engineering, and sports.

It has been almost 30 years since Simon and Chase published their first book on a general theory of expertise using research in the field of chess, and five years after that Elstein published the first book on clinical expertise and reasoning (Ericsson, 2007). Since then research on expertise in general and in the field of medicine has evolved dramatically. It is necessary to summarize the methods used to extract and study expertise, outline the stages of expertise development, present the practical implications of expertise research on medical curricula, and highlight the role of scheme application on enhancing expertise.

2.4.1 Extracting Expertise

Research in medical education studied clinical reasoning and expertise through three main domains, problem solving processes, memory performance, and mental representations (Norman, 2005). The knowledge structure and organization are eminent in expertise development. The study of knowledge structure has two methods, direct and indirect (McLaughlin et al., 2007). Indirect methods often employ statistical procedures to determine the knowledge required while giving insights into the relation between concepts/facts. Examples of indirect methods of investigation include multidimensional scaling, hierarchical clustering and concept sorting (McLaughlin et al., 2007). Direct methods assume that knowledge can be verbalized by the subject and appropriately classified by trained individuals to give insight to the

knowledge and knowledge structure. Examples of data gathering procedures include protocol analysis, interview and questionnaire.

2.4.1.1 Direct Techniques

Direct techniques assume that knowledge can be verbalized by the subject and appropriately classified by trained individuals to give insight to the subject's problem solving strategy and knowledge structure. The think aloud is a preferred method of inquiry at the initial stages to investigate the knowledge structure and problem-solving strategy within a domain, (Gilhooly, 1990; Patel, Groen, & Scott, 1988). A participant is presented with a realistic problem and solves it while verbalizing thoughts as much as possible. This method can yield useful information and does not disrupt normal processing (Ericsson & Simon, 1980).

2.4.1.2 Indirect Techniques

Indirect techniques often provide information on the relationship between concepts and make assumptions on what the candidate knows. Examples are concept sorting, multidimensional scaling, and hierarchical clustering. McLaughlin and colleagues in two different articles studied the medical knowledge structure of students using concept sorting tasks (McLaughlin & Mandin, 2002; McLaughlin et al., 2007). Participants of the study sorted important concepts into groups. From the sorting task, inferences of the hierarchical organizations of knowledge were made by the researchers. Concept sorting was found to be a reliable and valid technique to study medical knowledge structure.

Multidimensional scaling is similar to concept sorting but rather than sorting, items or concepts are rated for similarity or relatedness to each other (Olson & Rueter, 1987). The data are then statistically analyzed to determine the underlying knowledge structure.

2.4.2 Approaches to study Expert Characteristics

Chi (2006) described two approaches to study expertise; an absolute and a relative approach. The absolute approach involves studying distinguished exceptional people in their domain of expertise. The relative approach entitles studying experts in comparison to novices.

In the absolute approach experts have to be correctly identified first. Exceptional experts can be identified retrospectively with a quantitative index; for example, by number of publications in the domain of interest. A second way would be a rating system such as faculty ratings in a certain discipline or number of prestigious awards (e.g., Nobel). A third method would be to use an independent index or test of performance. In brief, when recognizing a distinguished expert, a reliable and valid method of measuring performance is required.

The relative approach involves a comparison between experts and novices. In this approach, experts are defined as relative to novices along a continuum (Hoffman, 1988). The advanced proficiency level can be assessed by years of experience, agreement among peers, or academic qualifications. The aim of studying relative expertise is to learn how expertise evolves and use that information to help learners/novices to become more skilled and knowledgeable (Chi, 2006). The relative approach to investigating expertise has been used in medical education to study the knowledge structure and clinical problem solving approach while comparing experts with novices (Coderre, Mandin, Harasym, & Fick, 2003; Daley, 1999; McLaughlin et al., 2007; McLaughlin & Mandin, 2002).

2.4.3 Evolution of Expertise

As the proficiency level advances, learners move along the continuum from novices to experts. A progression of knowledge structure and a change of the clinical problem solving methods utilized also occurs (Harasym, Tsai, & Hemmati, 2008). Attempts were made to formulate a theory of medical expertise on how medical knowledge evolves. The main goal of these studies was to identify the factors that distinguish experts from novices. Then these features can be utilized in training to enhance expertise. Research attempts were based on theories of process, memory or structure (Norman, 2005; Schmidt & Rikers, 2007). Process theories explained problem solving by analyzing the way doctors process patient-related information. Memory theories attempted to explain expertise in medicine by the extensive amounts of patient data recalled from a written clinical scenario and the advanced recall ability. Structure theories tried to relate expertise to the kind of knowledge and knowledge organization in memory. A summary of the preceding literature is provided in Appendix A.

2.4.3.1 Process Studies

Barrows and Abrahamson (1964) developed the technique of training actors to serve as SPs. This was considered a breakthrough in studying clinical problem solving (Elstein, Shulman, & Sprafka, 1978). Barrows and Bennet (1972) then studied the problem solving process of neurology residents using trained actors and concluded that problem solving was a process of early hypothesis generation with consequent testing of each hypothesis

Research on the process of clinical reasoning then followed by employing a direct technique of a think aloud protocol analysis. In 1968, Elstein, Shulman, and Sprafka (1990) began their problem solving studies at Michigan State University with high fidelity simulations. The group used a think aloud protocol during encounters of physicians with SPs. The main

finding was that all subjects initially generated a number of hypotheses and tested them deductively. This was a feature of all subjects regardless of their level of expertise. Another finding from this study was that problem solving is case specific and not a generalizable attribute.

At McMaster University, Neufeld et al. (1981) compared thoughts of experts and novices during videotaped interactions with SPs. Experts were experienced practitioners while medical students were less experienced subjects. Thinking process was studied by means of direct observation, videotape stimulated recall, and analysis of protocol transcripts. The main findings of early hypothesis generation, hypothetico-deductive reasoning, and case specificity were replicated and similar to the findings of Elstein and his colleagues.

Patel and Groen (1986) were unsatisfied with the previous findings that experts primarily generate and test hypothesis in the same way of novices. They used precise process models to clarify expert and novice differences in clinical reasoning and memory. Their empirical study was initiated by presenting participants with a written description of a case. Then they wrote as much information as possible that they recalled from the description (free-recall protocol). This was followed by a description of the underlying pathophysiology of the case. These two written think aloud (WTA) protocols were the basis for the propositional analysis. The propositions were considered by the authors as more precise models to access the clinical reasoning processes and knowledge base. The main finding was that physicians making an accurate diagnosis used forward reasoning, while the physicians diagnosed a case incorrectly generated and tested a number of hypotheses using backward reasoning. The authors were unsuccessful in detecting differences in recall of relevant patient information between forward-reasoning and backward-reasoning (Patel & Groen, 1986).

From the previous studies, it can be concluded that diagnostic accuracy was related to content knowledge and not to a general problem solving skill. This led investigators to shift research from expert problem-solving process to expert knowledge.

2.4.3.2 Memory Studies

Claessen and Boshuizen (1985) explained expert performance by superior knowledge and knowledge organization. They compared recall of case descriptions between students and doctors. The assumption was that meaningful patient information is recalled on the basis of pre-existing medical knowledge. In other words, the more detailed the content and knowledge structure of a physician the more relevant patient information was recalled.

Patel, Groen, and Frederiksen (1986) investigated students' and doctors' memory of clinical cases using propositional analysis. Two types of propositions were identified: recalled and inferred propositions. They found that novices recall more irrelevant details than experts. Experts remember more relevant medical information and were able to make meaningful inferences due to their highly developed knowledge base. They hypothesized that experts process the text at a deeper level.

In a study to better explain the role of memory measures in acquirement of expertise, Norman, Brooks, and Allen (1989) compared experts and novices in two experiments. The first experiment consisted of four written patient protocols with laboratory data. Subjects were deemed to be at three expertise levels; experts, intermediates, and novices. Each subject was given the patient protocol and asked to read it and then return the protocol to the examiner. Immediate verbal recall of the data and the reading time were recorded. The difference between the experts (specialized physicians) and the intermediates (students) was small and not significant with regards to the amount of data recalled. With this finding the authors suggested that intentional memory tasks are not suitable to study expertise in medicine.

In a second experiment, the researchers developed six protocols. Subjects were experts and intermediates. The major change from the previous experiment was in the instructions given to the participants. The first two protocols were given to participants and they were asked to describe how they formulate the problem and how the values relate to the problem formulation. After the second two protocols were discussed in the same manner, participants were asked if there are any laboratory values they can recall. This step was meant to test any incidental recall. The last two protocols participants were asked to read, memorize, and recall. The main finding of this experiment was that with incidental instructions, experts significantly recalled twice as much information as intermediates. The authors concluded that intentional memory is a misleading measure of expertise.

The above results were explained by Schmidt and Boshuizen (1993). That is, experts have an extensive knowledge base that is encapsulated until needed. When the task demands are increased by reducing time and increasing difficulty, expertise effects appear (Schmidt & Boshuizen, 1993).

2.4.3.3 Structure Studies

In an effort to explain the manner in which knowledge is organized in memory, Bordage, inspired by the work of Elstein in the 1970s, examined the application of prototype theory to medical knowledge (Rosch & Mervis, 1975). Accordingly, prototypes serve as anchors to new exemplars (i.e., typical or representative diseases). These are grouped into categories in memory. The central diseases are more common prototypical cases while the peripheral members/cases are less common diseases (Bordage & Zacks, 1984). This was a guiding teaching principle of Sir William Osler twenty five years ago, when the clerks working in his ward were directed to understand two disorders as a foundation, pneumonia and typhoid, and later experiences would build on these central diseases (Bordage, 2007; Bordage, 1987). The results of investigating the

applications of prototype theory provided evidence for the presence of prototypical organizations of disorders but could not explain the nature of the links between knowledge in memory (Bordage & Zacks, 1984).

Bordage (1994) then studied another theory to explain the nature of the relationships linking medical knowledge in memory. Structural semantic theory could provide a better understanding of the linkages between concepts in comparison to prototype theory. It implies that knowledge is given meaning through networks of relationships denoted by 2 different qualities, for example acute-chronic, local-systemic, and sharp-dull. Using qualitative think aloud protocol analysis, successful diagnosticians used more diversified and semantic qualifiers than less successful diagnosticians. This work on semantic theory led to proposing four categories of knowledge structure based on two dimensions, a semantic dimension founded by the number of semantic qualities used, and a syntactic dimension measured by the length of descriptions (Bordage, 1994).

Schmidt et al. (1990) described four developmental stages that lead to the development of four different knowledge structures. Harasym et al. (2008), inspired by the scholarly work of Bordage (1994) and Schmidt et al. (1990), developed a model that incorporates the knowledge structure and clinical problem solving approach. This model linked expertise process in the sense of the direction of reasoning with knowledge and knowledge structure. The knowledge structure evolves with deliberate practice and experience through the following stages: reduced, dispersed, elaborated causal, scheme/scripts.

In the stage of reduced knowledge, the learner has limited semantic content and few discourses. This embraces decreased knowledge of diseases and their manifestations. When knowledge of the diseases increases but with manifestations of each disease less developed, the

learner is in a dispersed knowledge stage and ultimately uses HD reasoning (i.e., testing of one hypothesis at a time) as a problem solving approach. There still is limited semantic content but with extended discourses (Bordage, 1994).

Further development and increase in knowledge of diseases and their manifestations with causal relationships, characterizes the elaborated causal stage. In this stage of an intermediate proficiency level, HD reasoning is used and better diagnostic accuracy is expected than the reduced stage. This is achieved with extensive and repeated application of acquired knowledge. This stage is characterized by richness of semantic content and extended discourses (Bordage, 1994; Schmidt et al., 1990).

The evolution of knowledge and knowledge structure into a hierarchical organized scheme with limited number of key features to discriminate each category is the scheme stage with semantic richness and limited discourses. This level of expertise is advanced and involves forward SI reasoning. It is associated with a high degree of diagnostic accuracy. With extended exposure to patients, an expert further establishes key features of diseases in a scheme, which enables recognition of diseases using pattern recognition (Schmidt et al., 1990; Bordage, 1994).

Deliberate feedback provides an opportunity to structure or restructure knowledge in learners' memory (Ericsson et al., 1993). Schmidt et al. (1990) stated that, as medical expertise highly relies on matching patient problems with previous encounters, then the number and structure of clinical problems in a curriculum is important. These theoretical studies in expertise and understanding of cognitive processes and knowledge structures have guided medical curriculum reform through two major structures (i.e., case and scheme based).

Ericsson is widely recognized as one of the world's leading theoretical and experimental researchers on expertise. His studies in various disciplines such as medicine, chess, and sports

investigate how expert performers acquire their superb performances. One of his conclusions was the "10,000-Hour Rule" – that is, it takes 10,000 hours of deliberate practice to be an expert in any field (Ericsson, Charness, Feltovich, & Hoffman, 2006).

Gladwell (2008) explains that reaching the 10,000-Hour Rule is simply a matter of practicing a specific task that can be accomplished with 20 hours of work a week for 10 years. It is believed that it takes approximately 10 years for an expert to develop a hierarchical knowledge structure within their speciality.

2.4.4 Practical Implications of the Studies on Expertise in Medicine

Papa and Harasym (1999) presented an in-depth review of curriculum reform in North America. The five major curricular models that have been developed to date include: the apprenticeship model; the disciplinary model; the organ-system based model; the problem/case based model; and the CP scheme-based model. Research to distinguish experts from novices regarding cognitive processes and knowledge representations were the steering influences for later stages of curricular reform. A summary of the educational practices for each model and its shortcomings are briefly discussed.

2.4.4.1 Apprenticeship Curriculum Model

The key educational practices identified in this model were memorizing lecture content and independent textbook study. Basic science was not integrated into clinical practice, and medical training was provided by a general practitioner that served as a mentor for one to three years. The main disadvantage of this model was that it had no defined competencies or standards to guide medical education.

2.4.4.2 Discipline Based Curriculum Model

The principles of this model were to develop thinkers that can generate and test hypotheses deductively, and to increase the amount of basic science knowledge taught. The main disadvantage of this model is that basic science disciplines were taught to a large extent with no integration between basic and clinical sciences.

2.4.4.3 Organ System Based Curriculum Model

In this model, an attempt was made to integrate the basic and clinical sciences within each organ system, strengthen students' intellectual (i.e., critical thinking and problem-solving) skills, and clearly define learning objectives (i.e., expected learning outcomes). The main disadvantages were that integration of knowledge did not improve the ability to differentially diagnose medical problems and lacked contextual learning.

2.4.4.4 Problem Based Curriculum Model

Instructional methods in this model were focused to develop students` HD learning in small group sessions, enhance active, self-directed learning, and improve differential diagnosis skills using medical problems. The main disadvantages were that problem solving was assumed to be a generic skill and a single disease is discussed for each CP (Elstein et al., 1978). Also, this model encourages backward reasoning, which is characteristic of insufficiently organized knowledge bases (Harasym et al., 2008).

2.4.4.5 CPC Model

This model aimed on centering the learning on the ways patients present to physicians, utilizing expert derived schemes to teach forward reasoning, and developing terminal and enabling objectives to integrate and balance basic and clinical sciences.

The University of Calgary was the originator and the first to apply the CPC model. One of the main features of the medical curriculum at the University of Calgary was the utilization of schemes (Mandin et al., 1995). These were developed by experts to help differentiate a cause from another within a CP and to guide problem solving. The CPC has been implemented for almost a decade with claims that scheme application helps organize students` knowledge bases for clerkships and advances expertise development (Mandin, Jones, Woloschuk, & Harasym, 1997).

2.4.5 Enhancing Medical Expertise through Scheme Application

The concept of schemes originates from the theoretical concept of prototypical knowledge structures or schemata. Glaser (1984) defines a schema as a modifiable information structure stored in memory. He recommends that novices use an organizational scheme as a scaffold to add new information and to provide basis for problem solving.

Kushniruk et al. (1998) from two experiments of studying experts and novices concluded that novices tend to generate large numbers of diagnostic hypotheses from different disease categories. Expert physicians on the other hand, rapidly select small sets of diagnostic hypotheses and use an efficient stepwise process to choose from the alternative hypotheses (Kushniruk, Patel, & Marley, 1998). These small sets of disease categories were labeled "small worlds" by the authors.

These theoretical concepts of "small worlds" or "prototypical knowledge structures" were utilized in the CPC to help students think like experts when solving clinical problems (Mandin et al., 1997). These "small worlds" in the context of the CPC were called schemes. The scheme was defined by Mandin et al. (1997) as "a mental categorization of knowledge that includes a particular organized way of understanding and responding to a complex situation". The aim was

to advance expertise development in medical students by providing them with schemes of CPs to assist the creation of knowledge structures and to guide scheme-based problem solving.

After the introduction of schemes to the medical school at the University of Calgary, students` perceptions on the utility of scheme-based problem solving was examined. Students reported schemes to be useful for learning and organizing information (Woloschuk, Harasym, Mandin, & Jones, 2000).

Coderre et al. (2003) analyzed think aloud protocols from experts and novices when attempting to diagnose four CPs. The aim was to provide empirical evidence and determine the relationship between problem-solving strategies and diagnostic success. They found that experts and novices utilized scheme-inductive reasoning had fivefold greater odds of diagnostic accuracy than those utilizing hypothetico-deductive reasoning. They concluded that the problem solving approach used, level of expertise, and CPs are significantly linked to diagnostic accuracy. This supports the notion of schemes as means of enhancing memory organization and improving diagnostic accuracy.

McLaughlin and colleagues in two different studies used concept sorting to investigate the relationship between knowledge organization and diagnostic success (McLaughlin & Mandin, 2002; McLaughlin et al., 2007). From both studies the use of schemes produced a positive outcome in performance. At the teaching level, scheme use by small group preceptors yielded increased number of expert type concepts in knowledge structure in first year medical students. The number of expert type concepts in knowledge structure was associated with increased odds of diagnostic accuracy.

Norman and Eva (2003) find these conclusions on the efficacy of scheme application to enhance diagnostic accuracy to be over simplified. The claim that the only way to prove that

there are inherent differences between two strategies, is to teach different strategies in a formal experiment then test their differential effects on problem solving. Experts and novices tend to use a combination of strategies depending on the problem, the stage and level of knowledge they are in with regards to the problem and solution, and their perception of what the experimenter wants to hear.

Coyle (2009) summarized the recipe for talent and expertise to evolve and that is practice, motivation and master coaching. It is not just practice or routine work that enhances expertise but deliberate practice with feedback improves specific aspects of an individuals` performance (Ericsson & Lehmann, 1996).

2.5 Problem Solving Strategies

Since the early studies on expertise and clinical problem solving, the notion of early generation and testing of hypotheses was repeated in literature (Barrows & Bennett, 1972; Neufeld et al., 1981). Experts' early hypotheses were found to be more diagnostically accurate. Experts tend to use contextual information more implicitly than novices and this leads to increased diagnostic accuracy (Hobu, Schmidt, Boshhuizen, & Patel, 2009). Contextual factors are illness features, such as age and gender, which favour the cause of a condition over another.

Curricular reform since 1765 to the present began slowly by building on strength of past innovations and overcoming weaknesses of past models (Papa & Harasym, 1999). Rapid changes in curricula occurred with the evolution of medical education research and understanding of cognitive processes and knowledge base structures.

One of the characteristics of experts is that they can display superior performance reliably on demand (Ericsson & Lehmann, 1996). The studies that investigated the differences between medical experts and non-experts sometimes were yielding contrasting information. For example, a study by Norman et al. (1999) concluded that the diagnostic performance for novices is more accurate with backward, diagnosis driven approaches. While Coderre et al. (2003) concluded that novices and experts achieved diagnostic success with scheme-induced forward reasoning. This discrepancy in findings could be due to the fact that researchers were unable to supply a valid measure with associated demonstrations of superior performance (Ericsson & Lehmann, 1996).

Studies of expertise highlighted some main features that distinguish experts from non-experts (Chi, 2006). Experts have acquired more knowledge in their domain, have organized and hierarchical structured knowledge, and can identify key-features that facilitate differentiating one cause of a disease from another. In medical education the basic elements defining expertise are knowledge, knowledge structure and processes of clinical reasoning (Schmidt et al., 1990).

As the proficiency level advances, a learner advances from a novice to an intermediate to an expert. A progression of knowledge structure and a change of the clinical problem solving methods utilized also occur (Harasym et al., 2008). As shown in Figure 2, the knowledge structure evolves with deliberate practice and experience through the following stages: reduced, dispersed, elaborated causal, scheme, and scripts. Depending on the knowledge structure, a corresponding problem solving strategy is utilized to clinically solve a problem. A detailed description of the problem solving strategy and its underlying knowledge structure is provided below.

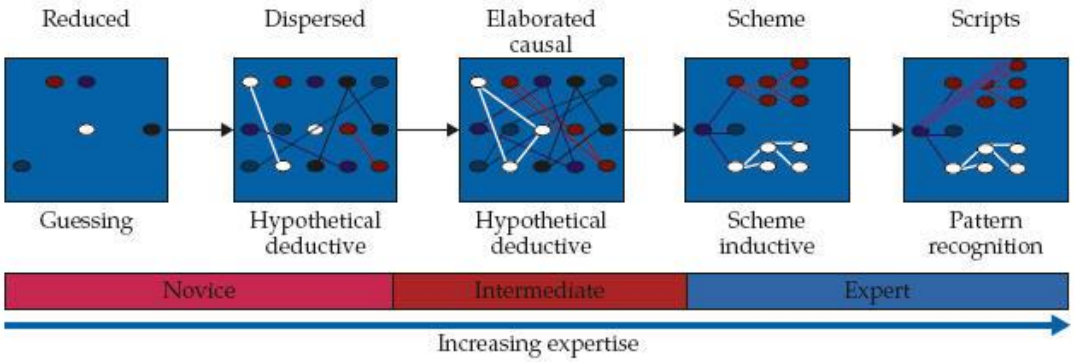


Figure 2. Evolution of knowledge structure from novice to intermediate to expert.

(Adopted from Harasym et al., 2008)

2.5.1 Guessing

In the stage of reduced knowledge, the learner has decreased knowledge of diseases and their manifestations. Therefore, the usual strategy used is guessing to solve a problem.

2.5.2 Hypothetical Deductive

When knowledge of the diseases increases but manifestations of each disease are less developed, the learner is in a dispersed knowledge stage and ultimately uses HD reasoning (i.e., testing of one hypothesis at a time) in clinical problem solving.

Further development and increase in knowledge of diseases and their manifestations with causal relationships characterizes the elaborated causal stage. In this stage of an intermediate proficiency level, HD reasoning is used and better diagnostic accuracy is expected than at the reduced stage.

2.5.3 Scheme Inductive

The evolution of knowledge and knowledge structure into a hierarchical organized scheme with a limited number of key features to discriminate each category is the scheme stage. This level of expertise is advanced and involves forward SI reasoning. It is associated with a high degree of diagnostic accuracy.

2.5.4 Pattern Recognition

With extended exposure to patients, an expert further establishes key features of diseases in a scheme, which enable recognition of diseases using pattern recognition. This advanced level of problem solving can only be achieved with extended exposure to clinical diseases. The diagnostician uses past experiences to recognize a pattern of clinical characteristics. These characteristics enable recognition of the clinical condition quickly.

2.6 Research Questions

Lack of research in predictability of diagnostic performance within CPs creates an opportunity to address the existence and degree of generalizability and the factors that may enhance performance. Clinical vignettes for diseases of two similar CPs were utilized to answer the following research questions:

1. *What is the underlying structure within clinical cases, and across two similar clinical presentations?*
2. *How consistent are diagnostic strategies across two cases in the same CP and across two similar CPs?*
3. *What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?*

Chapter Three: Research method

An exploratory factorial cohort design was conducted to answer the research questions outlined in chapter two. The purpose of this design was to explore the predictability of diagnostic performance within a CP, across two similar CPs, and to identify the effects of disease difficulty, case typicality, CP, and problem solving strategy on diagnostic performance. Section One describes the research participants; Section Two elaborates on the data sources and data collection method; Section Three presents the assumptions supporting the statistical analyses utilized to answer the three research questions; and Section Four explains the pertinent ethical issues in this research.

3.1 Research Participants

Participants were sixth year medical students enrolled in four universities in Saudi Arabia (SA). Most medical schools in SA enrol students after they finish high school in an undergraduate medical program. The main stream is a six year program. Students spend the initial three years in studying basic sciences and the latter three in clinical clerkships. A Bachelor of Medicine and Bachelor of Surgery (MBBS) are the two professional degrees granted after graduation. Although they are called two different degrees, they are awarded together as one. This follows the tradition of the British medical system. In countries that follow the tradition of North America, the equivalent postgraduate degree is the MD.

A convenience sample of final year medical students participated in this study. In total, 175 clerks were included: 71 from King AbdulAziz University in Jeddah; 41 from Taibah University in Madinah; 34 from Umm Al-Qura University in Makkah; and 29 from Qassim University in Buraidah.

3.2 Data Sources

Two types of data were collected; diagnostic performance from clinical vignettes and abstracted clinical problem solving strategies from WTA exercises. All participants from the four universities were invited to voluntarily participate in a 5 hour session in which medical clerks responded to 76 clinical vignettes and 4 WTA exercises. In return, each participant received feedback on diagnostic performance categorized by disease and CP.

3.2.1 Clinical Vignettes to Evaluate Diagnostic Performance

A clinical vignette includes a written case history, physical findings, and laboratory investigative results of a patient. It is based on a realistic clinical encounter and is followed by questions that investigate what a physician would do if presented with the actual patient (Veloski et al., 2005). Clinical vignettes have been used for more than two decades and are found to be a simple, economical, and valid indicator of a physicians' diagnostic and treatment competencies (Norcini, 2004; Peabody et al., 2004; Peabody et al., 2000).

In this study, 76 clinical vignettes were used to assess diagnostic performance. These clinical vignettes can be found in Appendix B1. Each clinical vignette was preceded by 19 different disease options taken from the two CPs used in this study. The disease options are listed in Appendix B2.

Only one single answer was correct for each clinical vignette. Participants were asked to indicate their answer by filling in the circle on an optical score sheet that corresponds to the letter of their answer (Appendix B3).

3.2.1.1 Clinical Presentations Used

Permission to utilize the University of Calgary Black Book schemes was granted by undergraduate medical education. Chest Discomfort and Dyspnea schemes were used to investigate diagnostic performance generalizability within and across CPs. Figure 3 below shows the cardiac causes of chest discomfort. The remainder CP schemes are provided in Appendix C1 to C5.

The two similar CPs were chosen because case specificity confines the generalizability of diagnostic performance from one CP to another. Also, the diseases within and across Chest Discomfort and Dyspnea CPs have a high degree of similarity in the initial signs and symptoms presentation. It was anticipated that participants will require a higher level of expertise to distinguish between atypical case presentations.

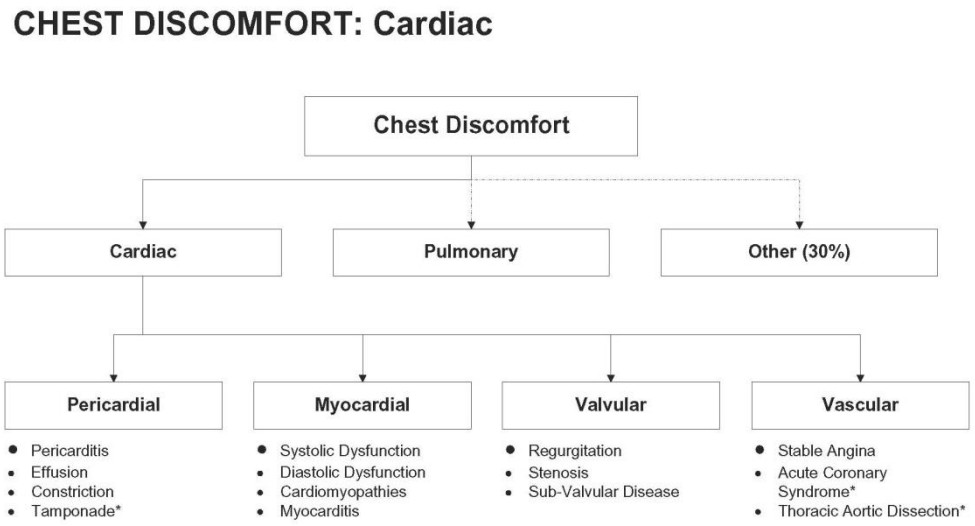


Figure 3. Cardiac Causes of Chest Discomfort

As shown in Table 1, nine cases from chest discomfort CP and ten from dyspnea were chosen to represent the various disease categories listed in the CP schemes found in Appendix C1 to C5.

Table 1. Diseases Included in the Study sorted by Clinical Presentation

Chest Discomfort	Dyspnea
Acute Coronary Syndrome	Acute Respiratory Distress Syndrome
Anxiety/Panic disorder	Anemia
Constrictive Pericarditis	Asthma
Lung Cancer	Cardiac Tamponade
Peptic Ulcer Disease	Chronic Obstructive Pulmonary Disease
Pneumonia	Congestive Heart Failure
Pneumothorax	Pulmonary Embolism
Stable Angina	Pulmonary Hypertension
Valvular Regurgitation	Sarcoidosis
	Valvular Stenosis

For each disease listed in Table 1, four clinical vignettes were created/selected/modified to represent the typicality gradient ranging from prototypical to atypical cases. In total 36 and 40 clinical vignettes were used to evaluate students' diagnostic performance of chest discomfort and dyspnea cases, respectively. Underground Clinical Vignettes series was the source for 64 clinical vignettes (Kim, Swanson, & Caplan, 2007). The question bank of the Saudi Council for Health Specialities (SCHS) was accessed with permission and the remaining 12 clinical vignettes suitable for this study were selected from their question bank.

3.2.1.2 Typicality Assessment and Validation of Clinical Vignettes

The first step of developing the test to evaluate diagnostic performance was to identify seventy six clinical vignettes that represent the nineteen cases used in this study. The next two

steps were to assess the typicality level and then validate the clinical vignettes. For these steps a cardiologist, pulmonologist, and a family physician, each with more than ten years of clinical experience, were the panel of experts identified to estimate typicality measures and face validity.

Typicality gradient is a continuum of patient presentations within a certain disease ranging from prototypical (usual signs and symptoms) to atypical (not usual signs and symptoms). Papa and Elieson (1993) defined case typicality as an estimate of the degree to which the case at hand both matches that disease's prototype, and, the degree to which it is dissimilar from its closest disease competitor.

In this study, a different method was used to identify case typicality. For each disease, the panel of experts were independently requested to evaluate the typicality level of each clinical vignette using a method that resembles abstraction theory. For example, for each of the four written cases for stable angina depending on the signs and symptoms presented, experts would define the percentage of clerks that would accurately diagnose the presentation provided. For each case, these were transformed into a range of one to four, one being most prototypical and four being most atypical.

Discrepancies in ratings were resolved through discussion and modifications made to adjust the clinical vignettes to have four levels of typicality for each disease. After an agreement by the expert panel was reached on the typicality ratings of seventy six clinical vignettes, the face validity of each case was reviewed and recommendations were considered to ensure proper representation of the clinical disease.

3.2.2 WTA Exercises to Identify Clinical Problem Solving Strategy

To evaluate the problem solving strategy used by participants in this study, a WTA protocol was utilized, with no time constraints. Participants were given four written clinical vignettes with four different diagnostic options to choose from for each case. The exercises can be found in Appendix D. After answering thirty clinical vignettes, two think aloud exercises were provided in the test package. The other two exercises were presented at the end of the 76 clinical vignette questions.

Clinical clerks were asked to describe, in writing, the thinking process they went through to answer the multiple-choice question and to justify the diagnosis chosen. The aim was to simplify the task and focus on documenting the thinking process without altering the result therefore, only four disease options were given with this exercise instead of nineteen. The classification by the two judges from each WTA vignettes were compared. Through discussion discrepancies in classification were removed. In addition, discrepancies in the two WTA for each CP were identified in order to generate on diagnostic strategy that was deemed to be used by each examinee in each CP.

3.3 Data Analyses

The following null hypotheses were assumed:

1-Diagnostic accuracy will not generalize within CP cases and across two similar CPs;

2-Problem solving strategies utilized will not be consistent from one case to another and across CPs;

3-There will be no difference in diagnostic performance by students using different diagnostic strategies.

The subsequent sections detail the assumptions underlying the statistical methods used to analyze the data collected.

3.3.1 Research Question 1

What is the underlying structure within clinical cases, and across two similar clinical presentations?

The aim was to examine the pattern of correlations between observed diagnostic performance scores. Items in the measures that are highly correlated (either positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

Principal Component Analysis (PCA) and Factor Analysis (FA) are two variable reduction strategies used to examine the relationship between and within CP/disease variables. Although, some researchers use PCA and FA interchangeably, they differ conceptually. As stated in O'Rourke, Hatcher, and Stepanski (2005, p 436.) "Both (PCA & FA) are methods that

can be used to identify groups of observed variables that tend to hang together empirically. Both procedures generally tend to provide similar results. Nonetheless, there are some important conceptual differences between PCA and FA that should be understood at the outset. Perhaps, the most important difference is with the assumption of an underlying causal structure. FA assumes that the co-variation in the observed variables is due to the presence of one or more latent variables (factors) that exert causal influence on these observed variables".

On the other hand, PCA reduces the number of observed variables to fewer principal *components,* which account for a *maximum amount of variance* of the observed variables. In contrast, FA analyzes the variance in the correlation structure using *common variance,* then hypothesizes underlying constructs (factors) – that is, variables not measured directly (Field, 2000). In the present study, PCA was employed as a variable reduction strategy to explain the variance and identify the groups of inter-related items or components.

The first step in the analyses involved conducting two separate PCAs for the nine cases in chest discomfort and ten cases of dyspnea. The purpose was to observe if the items within each CP group together in the same manner as they are grouped in the CP schemes – that is, by etiology, body system, or CP.

The second step was to determine if the 19 diseases cluster together into two CPs or sub-categories (i.e., body system or etiology), or if they cluster together under one group as a general diagnostic performance construct.

3.3.2 Research Question 2

How consistent are diagnostic strategies across two cases in the same CP and across two similar CPs?

Participants responded to four WTA exercises shown in Appendix D. For each CP, two clinical cases were randomly chosen. From the CP chest discomfort: stable angina and peptic ulcer disease and from dyspnea: COPD (chronic bronchitis) and asthma were selected. Each exercise consisted of a scenario with four disease options to choose from and only one was the correct diagnosis. After choosing the most appropriate answer, participants were asked to describe, in writing, the thinking process used to reach their selected answer. Four options were selected in order to minimize the length of time required to respond to each question while still providing adequate information to identify the diagnostic strategy used. The correct diagnosis was one of the options and the other three options were randomly selected from the 19 diseases used in the study. This WTA is a preferred method of inquiry at the initial stages to investigate the problem solving strategy utilized when diagnosing a clinical case and does not disrupt normal processing (Ericsson & Simon, 1980; Ericsson & Simon, 1980; Gilhooly, 1990).

Analyses of the written description involved classifying each thinking process into one of the four clinical problem solving strategies: guessing; HD; SI; or pattern recognition. The problem solving approach used by each participant was independently rated by two clinicians. For each CP, one clinical problem solving approach was identified. Discrepancies in decisions were resolved through discussion. The direction of problem solving and number of concepts related to the disease were the main factors used in categorizing the think aloud protocols.

Examples of written responses and corresponding diagnostic strategies are provided in Appendix E.

After determining the problem solving strategies used by each participant for chest discomfort and dyspnea, a cross-tabulation between the clinical problem solving categories was conducted. Cross-tabulation is a method used to analyze categorical variables.

To examine if the two problem solving strategies used in chest discomfort and dyspnea CPs are independent, Chi-square was computed. The Chi-square statistic is the primary statistic used for testing the statistical significance of the cross-tabulation table. If the variables are dependant, then the results of the statistical test are significant and the null hypothesis is rejected. This means that there is a relationship between the variables (Field, 2000).

3.3.3 Research Question 3

What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?

Multi-Facet Rasch Modelling (MFRM) is an analysis method that is based on Item Response Theory (IRT). MFRM generates a composite score based on the contribution of multiple facets and provides the percentage of variance removed from the true score (Bond & Fox, 2001).

MFRM was computed to determine the degree of impact (i.e., variance accounted for) by each of the independent variables. The assumption of uni-dimensionality for this statistical procedure was determined by PCA and fit statistics. The data were analyzed by FACETS (version 3.71.0) (Linacre, 2013). The program used the scores of diagnostic performance to estimate the effect disease difficulty, case typicality, and problem solving strategy on examinee abilities. FACETS calibrated diagnostic performance scores, disease difficulty, CP, case typicality, and problem solving strategies onto the same equal-interval scale (Linacre, 2005). This generated a logit-based scale for interpreting the results of the analyses.

3.3.3.1 IRT Assumptions

IRT is a probabilistic psychometric measurement model used to estimate ability and item characteristics on the trait measured (Bond & Fox, 2001). IRT presumes some assumptions that when understood help explain the theory (Bechger, Maris, Verstralen, & Berguin, 2003; Morris et al., 2006; Schaefer, 2008). These assumptions are:

1) One common factor accounts for all item covariances;

2) Relations between this common factor and the observed response take a particular form called an item characteristic curve;

3) All items must contribute in a meaningful way to the attribute being investigated.

The first assumption is that one common factor accounts for all item covariances. This single common factor is the latent trait being measured. The assumption is presented in literature as uni-dimensionality and local independence (Fan, 1998). Uni-dimensionality assumes that one attribute or dimension is being measured called a single latent trait. Confusing a number of attributes into a single score makes the score a less useful measure of that attribute of ability.

Local independence requires that the success or failure on an item is not dependent on the success or failure on any other item (Baghaei, 2008). Local independence includes uni-dimensionality but goes beyond it. Including the same item twice does not affect the dimensionality of a test. Responses to the two identical items would not be locally independent as the test taker would either succeed or fail on both items (Wright, 1996).

This first assumption is never met precisely. It is obviously a problem when a test includes multiple items that are related by a common construct. MacDonald and Pantone (2002) suggest that minor violations of this assumption do not make a difference as long as a clear first dominant factor is identified. Wright and Linacre (1989) argue that no test can be perfectly uni-

dimensional and that uni-dimensionality is conceptual and qualitative rather than factual and quantitative. So, uni-dimensionality and local item independence are always a matter of assumption. It is possible, however, to validate these assumptions; this will be discussed later in this section.

The second assumption states that relations between the latent trait and the observed response take the form of an item characteristic curve (ICC). ICC plots the probability of responding correctly to an item as a function of the latent trait under investigation (Crocker & Algina, 2006).

Commonly, in IRT, ICC takes an S-shape. As the score on the latent trait increases, the probability to answer an item correctly increases. The importance of ICC is that it shows how the probability of answering correctly depends on the latent trait. It is important to properly interpret the probability of responding correctly to an item from the ICC.

Lord (1980) suggested it is necessary not to interpret the probability of responding correctly as the probability of a specific person answering a specific item correctly. The proper interpretation would be theorizing a subgroup of persons or examinees at each point on the latent trait scale that have the same latent trait score. The probability of answering correctly is the probability that a randomly chosen person of this subgroup will respond correctly to an item. For example an ICC with the latent trait score of 3, and the probability of responding correctly 0.45. This can be interpreted that a randomly chosen person with a latent score of 3, has the probability of 0.45 to answer correctly.

The third assumption is that all items must contribute in a meaningful way to the attribute being investigated. A proper measurement model will only allow for estimation of one attribute at a time. This assumption of IRT focuses on uni-dimensional construct validity, in which the scores or abilities are reflecting a single underlying construct (Bond & Fox, 2001). This helps in validating the assumption of uni-dimensionality and works as a quality control mechanism. IRT analyses provide fit statistics that are indicators of how well each item fits within the underlying attribute or construct. Fit statistics are designed to flag statistically unexpected responses or response patterns by considering the performance of one respondent or one test item at a time.

There are three main approaches in assessing the dimensionality of an examination: 1) using PCA only, 2) using fit statistics only, and 3) using fit statistics, then exposing the residuals to a PCA (Tennant & Pallant, 2006).

The first approach, PCA of the original observations, usually reports items with different difficulty levels as different factors (Bond, 1994). This is also subjected to error as original observations are non-linear. Linacre suggested using the third approach, fit statistics then running a PCA of residuals (Linacre, 1998). IRT analysis converts the original observations into an interval variable. The dominant variable/dimension may be a mix of items measuring communication-examination competency; lesser dimensions are reported as misfit. A PCA of residuals would separate the communication-examination items.

3.3.3.2 Sample Size

Hambleton and Jones (1993) explain that IRT models require larger sample sizes than classical test theory (CTT) (Hambleton & Jones, 1993). While CTT requires 200 to 500 examinees to accurately run an analysis, Downing (2003) notes the requirement of sizable samples of examinees for IRT. A 3-parameter IRT model that includes difficulty, discrimination, and guessing, requires more than 1000 examinees.

For a 1-parameter IRT model, which includes difficulty only, a sample size of 30 for robust decision making is required (Linacre, 1994). Rasch modelling is considered a 1-parameter IRT model. In the current study, a form of Rasch modelling, MFRM, is computed therefore the sample size of 175 participants was deemed sufficient.

Table 2. Method Summary

Research Question	Variables	Statistical Analyses
What is the underlying structure within clinical cases, and across two similar clinical presentations?	Total score from 4 clinical vignettes on each of the 19 diseases.	PCA
How consistent are diagnostic strategies across two cases in the same CP and across two similar CPs?	Categorical Variable: Clinical problem solving approaches on each CP.	CROSS-TABULATION CHI SQUARE
What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?	Dependant: Score on each clinical (total 76). Independent: disease, typicality, type of CP, clinical problem solving strategy for each CP.	MFRM

3.4 Ethical Considerations

This study received ethical approval from the Institutional Review Board (IRB) at King Fahd Medical City and from the University of Calgary`s Conjoint Health Research Ethics Board (CHREB). Each participant reviewed and signed a consent form. The informed consent provided basic information of what the research is about and what his/her participation involved. It clearly stated that participation in this study was completely voluntary and may be withdrawn at any time. All data collected had a unique numerical identifier. The data were placed anonymously in a folder after completion. The researcher in charge of the study had access to the information as a failsafe; however no information was disclosed to anyone else.

Chapter Four: Results

4.1 Overview

The goal of the current study was to examine the predictability of diagnostic accuracy and the factors that have an effect on diagnosis in two similar CPs. The research goals were to investigate the underlying structure within clinical cases and across CPs; study the effect of problem solving approach on diagnostic accuracy; and examine the effect of facets (disease, typicality, CP, and problem solving approach) on diagnostic accuracy. This chapter presents the results of the statistical analyses used to answer the research questions.

4.2 Descriptive Data

Socio-demographic data of 175 clerks and participating universities are summarized below in Table 3. Three of the four universities were located in the western region of SA. In total, the current study involved more female than male clerks (66% vs. 34%).

Table 3. Descriptive Data of Participants by University

University	City/Region	Female (%)	Male (%)	Total # of Participants
King AbdulAziz	Jeddah/Western	61 (86)	10 (14)	71
Taibah	Madinah/Western	22 (54)	19 (46)	41
Umm AlQura	Makkah/Western	24 (71)	10 (29)	34
Qassim	Qassim/Central	9 (31)	20 (69)	29
Total		116 (66)	59 (34)	175

4.3 Item Reliability

Chest Discomfort CP consisted of 36 items (α = 0.88), Dyspnea CP consisted of 40 items (α = 0.84). When scores from both CP items were compiled into the representative 19 diseases, the estimate of internal consistency Cronbach's alpha was 0.91, indicating high reliability.

4.4 Results related to Research Question 1

What is the underlying structure within clinical cases, and across two similar clinical presentations?

To determine the underlying structure of the clinical cases within a CP and across two similar CPs, three separate PCAs were employed. Results are presented in the following sections.

4.4.1 PCA for Chest Discomfort Cases

The nine clinical cases of chest discomfort are shown in Figure 4. PCA was conducted to examine the level at which cases group together. Clustering can occur at levels of CP, body systems, or organ/tissue structures.

4.4.1.1 Correlation

The starting point for all variable reduction techniques is the correlation matrix. All FA techniques attempt to cluster subgroups of variables based on their correlations. PCA requires that there be some correlations greater than 0.30 between the variables included in the analysis.

Correlation analyses were used to examine the relationship between clerks' diagnostic accuracy on nine chest discomfort clinical cases. The results in Table 4 demonstrate that 34 out of 35 two-tailed correlations were statistically significant and ranged from $r(173) = +0.20$, $p < 0.01$, to $r(173) = +0.57$, $p < 0.01$. The correlations of diagnostic performance scores in constrictive pericarditis with scores from lung cancer were not significant. Overall the results

suggest that there is a relationship between diagnostic performance scores of cases within chest discomfort CP.

4.4.1.2 Principal Components Analysis

PCA was conducted as part of a variable reduction approach to examine construct validity and to answer question 1 by identifying the groups of inter-related diseases within chest discomfort CP. PCA requires that the probability associated with Bartlett's Test of Sphericity be less than the level of significance. In the current study the probability associated with the Bartlett test is < 0.001, which satisfies this requirement. In addition, the overall Measuring of Sampling Adequacy (MSA) for the set of variables included in the analysis was 0.89, which is high and exceeds the minimum requirement of 0.50 for overall MSA.

PCA for nine chest discomfort cases yielded one component with an eigenvalue of 4.15. This component described 46.15% of the total variance. Component one was labeled chest discomfort. Table 5 summarizes the PCA findings for chest discomfort cases.

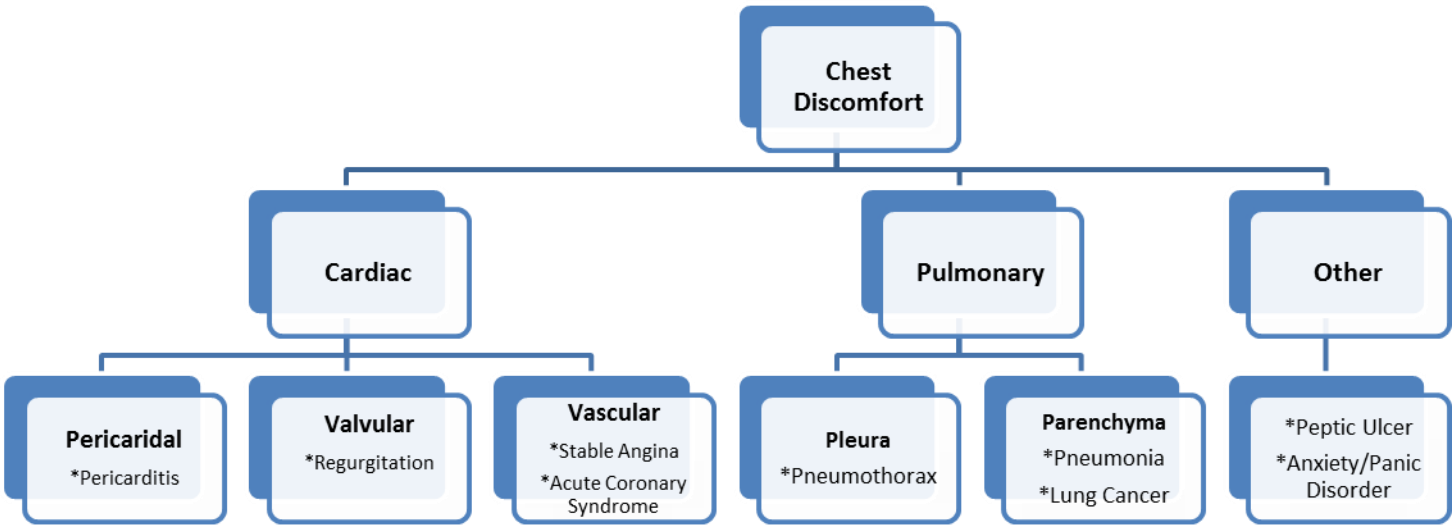


Figure 4. Chest Discomfort Scheme for Cases used in Study

*Clinical Case

Table 4. Chest Discomfort cases inter-correlations and descriptive statistics

	ACS	ANX	CONSTPERI	LUNGCAN	PEPULC	PNEUMONIA	PNEMTHX	STABANGI	VALVREG	MEAN	SD
ACS	1.00									2.98	1.09
ANX	0.37**	1.00								3.18	.91
CONSTPERI	0.33**	0.20**	1.00							1.74	.97
LUNGCAN	0.44**	0.43**	0.14	1.00						3.03	1.01
PEPULC	0.57**	0.35**	0.20**	0.47**	1.00					3.37	.81
PNEUMONIA	0.61**	0.42**	0.30**	0.56**	0.52**	1.00				3.24	.99
PNEMTHX	0.46**	0.38**	0.34**	0.46**	0.39**	0.49**	1.00			2.90	1.12
STABANGI	0.48**	0.37**	0.20**	0.32**	0.39**	0.42**	0.39**	1.00		2.67	1.10
VALVREG	0.43**	0.28**	0.31**	0.28**	0.36**	0.46**	0.42**	0.36**	1.00	1.80	1.14

**Correlation is significant at the 0.01 level (2-tailed).

Table 5. Principal Component Analysis for Chest Discomfort Cases

Item	Disease	Communality	Component
			Chest Discomfort
ACS	Acute Coronary Syndrome	0.62	0.78
ANX	Anxiety/Panic Disorder	0.38	0.62
CONSTPERI	Constrictive Pericarditis	0.21	0.45
LUNGCAN	Lung Cancer	0.47	0.68
PEPULC	Pulmonary Embolism	0.51	0.71
PNEUMONIA	Pneumonia	0.64	0.80
PNEUMOTHARAX	Pneumothorax	0.51	0.71
STABANGI	Stable Angina	0.41	0.64
VALVREG	Valvular Regurgitation	0.41	0.64
# of Diseases	9		
Cronbach's α	0.88		
Kaiser-Meyer-Olkin (KMO)	0.89		
Bartlett's test of Sphericity	0.000		
Total variance explained	46.15%		
Eigenvalue	4.15		

4.4.2 PCA for Dyspnea Cases

The ten clinical cases of dyspnea are shown in Figure 5. PCA was conducted to examine the level at which cases group together. Clustering can occur at the following levels: CP; body systems; or organ/tissue structures.

4.4.2.1 Correlation

Correlation analyses were used to examine similarity in clerks' diagnostic accuracy across 10 dyspnea clinical diseases. The results in Table 6 show that 41 out of 44 two-tailed correlations were statistically significant and ranged from $r(173) = +0.17$, $p < 0.05$, to $r(173) = +0.55$, $p < 0.05$. The correlations of diagnostic performance scores in cardiac tamponade with scores from valvular stenosis; asthma; and COPD were not significant. Overall, the results suggest that there is a relationship between diagnostic performance scores of cases within dyspnea CP except cardiac tamponade, which had either low or non-significant correlations with other cases.

4.4.2.2 Principal Components Analysis

PCA was conducted as part of a variable reduction approach to examine construct validity and to answer question 1 by identifying the groups of inter-related diseases within the CP of dyspnea. In the current analyses the probability associated with the Bartlett test is < 0.001. The overall Measuring of Sampling Adequacy (MSA) for the set of variables included in the analysis was 0.86, which is high and exceeds the minimum requirement of 0.50 for overall MSA.

First, PCA with ten dyspnea cases was performed. The analysis yielded two components following rotation to normalized Varimax and Kasier criteria with eigenvalues 4.09 and 1.02. These components described 51.2% of the total variance. Component one had eight items loading on it. Cardiac tamponade loaded on component two. The remaining item cross-loaded between components. Due to the fact that the cardiac tamponade showed low or no significant

correlations with other cases and solely loaded on one component; this item was dropped from the PCA.

Secondly, PCA with nine dyspnea cases excluding cardiac tamponade was performed. As shown in Table 7, PCA for nine dyspnea cases yielded one component with an eigenvalue of 4.01. This component described 44.60% of the total variance. Component one was labeled dyspnea.

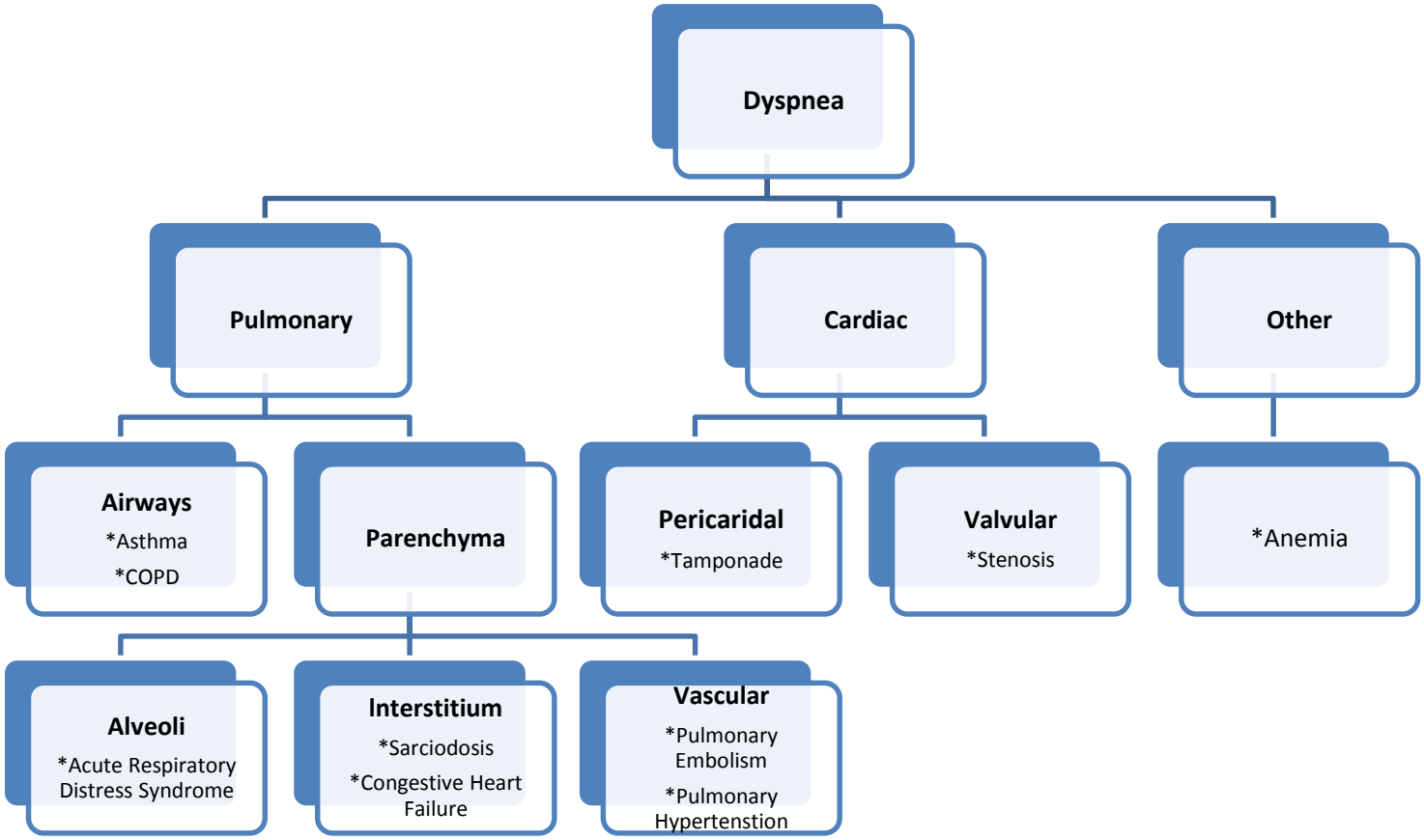


Figure 5. Dyspnea Scheme for Cases used in Study

*Clinical Case

Table 6. Dyspnea cases inter-correlations and descriptive statistics

	ASTHMA	COPD	PULMEMBO	CARDTAMP	ANEMIA	ARDS	CHF	PULMHYP	VALVSTEN	SARCOID	MEAN	SD
											2.46	0.91
ASTHMA	1.00										2.75	1.13
COPD	0.36**	1.00									2.90	1.13
PULMEMBO	0.30**	0.51**	1.00								1.78	0.95
CARDTAMP	0.06	0.09	0.21**	1.00							3.01	1.00
ANEMIA	0.37**	0.39**	0.56**	0.22**	1.00						1.15	1.03
ARDS	0.18*	0.28**	0.29**	0.18**	0.31**	1.00					2.26	0.80
CHF	0.34**	0.40**	0.52**	0.23**	0.28**	0.25**	1.00				1.45	1.02
PULMHYP	0.31**	0.39**	0.40**	0.17*	0.39**	0.40**	0.31**	1.00			2.56	1.24
VALVSTEN	0.21**	0.45**	0.55**	0.09	0.31**	0.30**	0.48**	0.24**			2.89	1.18
SARCOID	0.19*	0.41**	0.60**	0.23**	0.39**	0.40**	0.41**	0.32**	0.50**	1.00	2.46	0.91

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Table 7. Principal Component Analysis for Dyspnea Cases

Item	Disease	Communality	Component
			Dyspnea
ASTHMA	Asthma	0.26	0.51
COPD	Chronic Obstructive Pulmonary Disease	0.50	0.71
PULMEMBO	Pulmonary Embolism	0.67	0.82
ANEMIA	Anemia	0.45	0.67
ARDS	Acute Respiratory Distress Syndrome	0.30	0.54
CHF	Chronic Heart Failure	0.45	0.67
PULMHYP	Pulmonary Hypertension	0.38	0.61
VALVSTEN	Valvular Stenosis	0.48	0.69
SARCOID	Sarcoidosis	0.52	0.72
# of Diseases	9		
Cronbach's α	0.84		
Kaiser-Meyer-Olkin (KMO)	0.86		
Bartlett's test of Sphericity	0.000		
Total variance explained	44.60%		
Eigenvalue	4.01		

4.4.3 PCA for Chest Discomfort and Dyspnea Cases

4.4.3.1 Correlation

Correlation analyses were used to examine the relationship between clerks' diagnostic performance scores in 19 clinical diseases representing chest discomfort and dyspnea CPs. The results in Table 8 suggest that 163 out of 171 two-tailed correlations were statistically significant and ranged from $r(173) = +0.17$, $p < 0.05$, to $r(173) = +0.66$, $p < 0.05$. The correlations of diagnostic accuracy in cardiac tamponade with other diseases were not significant in 6 out of 18 correlations. Performance score correlations in constrictive pericarditis were not significant with lung cancer $r(173) = +0.14$, $p > 0.05$ and asthma $r(173) = +0.02$, $p > 0.05$. In general, the results suggest that there is a relationship between diagnostic performances in most of the diseases within the two similar CPs, with the exception of cardiac tamponade and constrictive pericarditis.

4.4.4 Principal Components Analysis

PCA was conducted as part of a variable reduction approach to examine construct validity and to answer question 1 by identifying the groups of inter-related diseases. PCA requires that the probability associated with Bartlett's Test of Sphericity be less than the level of significance. The probability associated with the Bartlett test is < 0.001, which satisfies this requirement. In addition, the overall Measuring of Sampling Adequacy (MSA) for the set of variables included in the analysis was 0.92, which is high and exceeds the minimum requirement of 0.50 for overall MSA.

First, an unforced PCA solution was performed. The analysis yielded three components following rotation to normalized Varimax and Kasier criteria with eigenvalues 5.12, 3.19, and 2.10, respectively. These components described 54.87% of

the total variance. Components one to three had nine, three, and two items loading on them, respectively. The remaining five items cross-loaded between components. Component one and two could not be labeled distinctively, as loaded items represented a combination of diseases from both chest discomfort and dyspnea. The two items loading on component three were the diseases that least correlated with others in the initial correlation analyses. These two items combined, cardiac tamponade and constrictive pericarditis are grouped as pericardial causes of chest discomfort.

Due to the undefined component loadings and the fact that two types of cases were originally assessed in the questions (i.e., chest discomfort and dyspnea); a forced two-factor PCA solution was conducted. As shown in Table 9, two components were identified, following rotation to the normalized Varimax and Kaiser method with eigenvalues equal to 6.24 and 2.83, respectively. The two components explained 47.71% of the total variance.

In interpreting the rotated components, an item was identified under a given component if the loading was 0.50 or greater for that component and was less than 0.50 for the other. Using these criteria, 13 out of 19 items loaded on the first component, which was subsequently labelled CPs. The 4 items loading on component two are the most difficult diseases according to MFRM analyses. The second component was labelled most difficult diseases. Two items showed significant cross loadings between components. This PCA finding validates the uni-dimensionality assumption, as component two is not a secondary dimension but rather a component that resulted from the four most difficult items.

Table 8. Chest Discomfort and Dyspnea cases inter-correlations and descriptive statistics

	CONST-PERI	lung can	asthma	copd	pulm embo	card tamp	anemia	PEP ULC	ANX	ards	CHF	pulm hyp	VALV REG	valv sten	ACS	STAB ANGI	sarco	PNEUM-ONIA	PNEUM-THX	MEAN	SD
CONST PERI	1.00																			1.74	.97
lung can	.14	1.00																		3.03	1.01
asthma	.02	.37**	1.00																	2.46	.91
copd	.17*	.45**	.36**	1.00																2.75	1.13
pulm embo	.32**	.58**	.30**	.51**	1.00															2.90	1.13
card tamp	.36**	.05	.06	.09	.21**	1.00														1.78	.95
anemia	.22**	.50**	.37**	.39**	.56**	.22**	1.00													3.01	1.00
PEP ULC	.20**	.47**	.17**	.39**	.58**	.12	.39**	1.00												3.37	.81
ANX	.20**	.43**	.27**	.32**	.44**	.21**	.52**	.35**	1.00											3.18	.91
ards	.24**	.23**	.18**	.28**	.29**	.18*	.31**	.17*	.28**	1.00										1.15	1.03
CHF	.33**	.48**	.34**	.40**	.52**	.23**	.28**	.50**	.39**	.25**	1.00									2.26	.80
pulm hyp	.34**	.22**	.31**	.39**	.40**	.17*	.39**	.21**	.20**	.40**	.31**	1.00								1.45	1.02
VALV REG	.31**	.28**	.15**	.38**	.40**	.25**	.28**	.36**	.28**	.29**	.40**	.24**	1.00							1.80	1.14
valv sten	.29**	.43**	.21**	.45**	.55**	.09	.31**	.61**	.37**	.30**	.48**	.24**	.53**	1.00						2.56	1.24
ACS	.33**	.44**	.18**	.36**	.63**	.14	.35**	.57**	.37**	.18*	.49**	.25**	.43**	.58**	1.00					2.98	1.09
STAB ANGI	.20**	.32**	.16**	.30**	.40**	.24**	.30**	.39**	.37**	.19*	.42**	.17*	.36**	.46**	.48**	1.00				2.67	1.10
sarco	.33**	.52**	.19**	.41**	.60**	.23**	.39**	.43**	.43**	.40**	.41**	.32**	.43**	.50**	.49**	.32**	1.00			2.89	1.18
PNEUMONIA	.30**	.56**	.34**	.46**	.66**	.17*	.41**	.52**	.42**	.23**	.49**	.33**	.46**	.52**	.61**	.42**	.45**	1.00		3.24	.99
PNEUMTHX	.34**	.46**	.33**	.48**	.48**	.31**	.44**	.39**	.38**	.41**	.46**	.43**	.42**	.45**	.46**	.39**	.49**	.49**	1.00	2.90	1.12

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Upper Case: Chest Discomfort Cases

Lower Case: Dyspnea Cases

Table 9. Forced 2-Factor Solution Principal Component Analysis

Item	Disease	Communality	Rotated Components	
			CP	Difficult Diseases[1]
PEPULC	PEPTIC ULCER	0.58	**0.76**	0.02
ANX	ANXIETY	0.37	**0.54**	0.27
CHF	CHRONIC HEART FAILURE	0.48	**0.64**	0.27
VALVREG	VALVULAR REGURGITATION	0.38	**0.50**	0.36
ACS	ACUTE CORONARY SYNDROME	0.58	**0.75**	0.13
STABANGI	STABLE ANGINA	0.35	**0.56**	0.17
PNEUMONIA	PNEUMONIA	0.61	**0.75**	0.20
PNEUMOTHARAX	PNEUMOTHORAX	0.57	0.52	0.54
CONSTPERI	CONSTRICTIVE PERICARDITIS	0.44	0.16	**0.64**
copd	chronic obstructive pulmonary disease	0.42	**0.59**	0.26
pulmembo	pulmonary embolism	0.67	**0.77**	0.27
anemia	anemia	0.41	**0.51**	0.38
sarcoid	sarcoidosis	0.51	**0.59**	0.39
lungcan	lung cancer	0.55	**0.73**	0.08
cardtamp	cardiac tamponade	0.42	0.01	**0.65**
ards	acute respiratory distress syndrome	0.43	0.18	**0.62**
pulmhyp	pulmonary hypertension	0.46	0.23	**0.64**
valvsten	valvular stenosis	0.63	**0.74**	0.14
asthma	asthma	0.18	0.39	0.18
# of Diseases	19		13	4
Cronbach's α	0.91	Cross Loading	2	
Kaiser-Meyer-Olkin (KMO)	0.92			
Bartlett's test of Sphericity	0.000			
Total variance explained	47.71%		32.84%	14.88%
		Eigenvalue	6.24	2.83

*Diseases[1] = 4 difficult diseases clustered together likely due to reduced variance in clerks' diagnostic scores.

Upper Case: Chest Discomfort Cases Lower Case: Dyspnea Cases

4.5 Results related to Research Question 2

How consistent are diagnostic strategies across two cases in the same CP and across two similar CPs?

In this study, the three categorical values are guessing, HD, SI. Only three cases were identified as pattern recognition and were collapsed into the category of scheme induction. To determine if a relationship exists in reasoning strategies between two similar CPs, cross-tabulation was examined by the chi-square test of association. As demonstrated in Table 10, 56.5% participants used guessing to solve CD and DYS cases and 43.5% of those you used guessing on CD cases used HD reasoning on the DYS cases. HD reasoning was used by 83.30% of the participants in both CD and DYS cases. Of those who used HD in the CD cases, 8.7% used guessing and 7.9% used SI for solving the DYS cases. As for SI problem solving strategy, 50% of participants used the same approach in both CPs. In addition, 50% of those who used SI in DYS used HD in CD. None of the clerks used guessing in a CP while using scheme induction in the other.

There was a statistically significant relationship between problem solving strategies used by clerks in this study to diagnose chest discomfort and dyspnea cases, as the Chi-square results show that $\chi^2(1, N = 175) = 75.53, p < 0.001$.

Table 10. Cross tabulation of Problem Solving Strategies used for Chest Discomfort and Dyspnea

			Problem Solving Strategy Dyspnea (DYS)			
			Guessing DYS	Hypothetical Deductive DYS	Scheme Inductive DYS	Total
Problem Solving Strategy Chest Discomfort (CD)	Guessing CD	Count	13	10	0	23
		% within Approach CD	**56.5%**	43.5%	0.0%	100.0%
	Hypothetical Deductive CD	Count	11	105	10	126
		% within Approach CD	8.7%	**83.3%**	7.9%	100.0%
	Scheme Inductive CD	Count	0	13	13	26
		% within Approach CD	0.0%	50.0%	**50.0%**	100.0%
Total		Count	24	128	23	175
		% within Approach CD	13.7%	73.1%	13.1%	100.0%

4.6 Results related to Research Question 3

What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?

To study the effect of disease difficulty, case typicality, type of CP, and problem solving strategy on diagnostic performance - that is, candidate ability, MFRM was conducted using FACETS software.

MFRM was computed with 175 candidates, 76 disease items, four typicality levels, two CPs, and three problem solving strategies. The variance explained by the Rasch measures was only 24.15%. This variance-explained is dependent on the variances by the measures of the elements in the facets.

All the facets have variances smaller than one logit because the standard deviations were low. Therefore, variance explained by Rasch measures can only be small (Linacre, 2005). This finding of low standard deviations was attributed to the fact that the 76 cases were assigned scores ranging from zero to one. To adjust for this low variance, the observed scores were re-grouped. Performance scores on the four clinical vignettes that represent each disease were summed. This led to observed scores ranging from zero to four for each of the 19 diseases. In addition, the scores for typicality and diagnostic strategy were summed to reflect a total score for each of the 19 diseases.

4.6.1 Model Fit and Uni-Dimensionality

MFRM analysis was run and nine disjoint subsets in the CP facet were detected. Subsets are found when the elements of the facets are not sufficiently crossed. In the current data,

diseases are nested within case typicality and CP. To resolve the subset connections, case typicality and CP were converted into “demographic" facets. Demographic facets are not included in the measurement model. After the adjustment of the model, subset connection in the data was achieved with the measurement model including candidates, disease, and problem solving approach. Thus, the estimates of typicality and CP were excluded from the analysis.

Model fit and uni-dimensionality were estimated using standard residuals and infit-outfit indices. Standardized residuals mean was 0.01 and the sample standard deviation was 1.00. These results indicate that the data fit the Rasch model.

Uni-dimensionality of data is a MFRM assumption. Fit statistics revealed four candidates (2.28%) out of 175 with infit Z standardized statistics (Zstd) that were less than the critical value of -2 and ten (5.71%) with outfit Zstd statistics that were greater than the critical criteria of +2.0 (Bond & Fox, 2001). Of the 19 disease items, one (5.26%) had infit Zstd statistics less than -2 and one (5.26%) had Zstd statistics greater than the criterion of +2.0. The percentage of candidates and diseases showing acceptable fit statistics did not fall below 90%. These findings with the PCA conducted in research question one support the uni-dimensionality of the data.

4.6.2 Effect of Facets

Reliability estimates were estimated for each facet and the gold standard was set at 0.80. The estimated reliabilities for the facets candidates, diseases, and problem solving approach are 0.86; 0.99; and 0.99, respectively. This indicates high reliability.

MFRM analysis revealed that 56.80% of the main effects variance was explained by Rasch-measures. Of the total systematic variance, candidate ability accounted for 17.74.8%, differences in disease difficulty resulted in 29.71% variance, and 9.34% was due to problem solving strategy.

4.6.3 Candidate Ability

The abilities of 175 clinical clerks ranged from +192 till -131 logits. According to Rasch estimated true diagnostic ability, candidate 102 had the highest ability of +192 logits. The raw score of this candidate was 68 out of 76. In contrast, candidate 111 had a raw score 15 out of 76, and had an ability Rasch true measure of -131. Appendix F shows the candidates` measurement report.

4.6.4 Difficulty and Discrimination Indices of Diseases

Disease difficulties are listed in Appendix G. Diseases varied in difficulty with a range of +148 to -91 logits. The most difficult disease to diagnose was acute respiratory distress (total score 202 and logit score of +147) while the easiest was peptic ulcer disease (total score 589 and logit score of -91).

Without altering other estimates, an estimate of item discrimination is computed in MFRM (Linacre, 2013). Disease discrimination indices for 14 out of 19 items were equal or close to the desired Rasch expectation of $di = 1$.

4.6.5 Clinical Problem Solving Approach

Clinical problem solving approach MFRM measurements are reported in Appendix E. SI reasoning was the most discriminating with a of $di = 1.14$. Clerks who used guessing had a total score of 637 (+56 logits), those who used HD reasoning had a total score of 6288 (-1

logits), and students who used scheme induction had a total score of 1492 (-34 logits). This finding indicates that clerks who tended to use guessing as a diagnostic strategy were least likely to obtain the correct diagnosis (mean score of 26.64), clerks who tended to use HD reasoning were more likely to get the correct diagnosis (mean score of 50.31), and clerks who used scheme induction were most likely to answer correctly (mean score of 62.38). In both CPs, there was a statistically significant difference among the three strategy groups as determined by one-way ANOVA, $F(2,172) = 85.19$, $p < 0.001$ for chest discomfort cases and, $F(2, 172) = 84.87$, $p < 0.001$ for dyspnea cases.

Chapter Five: Discussion

Diagnostic performance measures for 175 clerks were assessed using 76 clinical vignettes. These vignettes represented 19 diseases in two CPs: chest discomfort or dyspnea. The data were analyzed using quantitative statistical methods to identify the underlying structure across the two CPs and within the 19 diseases; to study the effect of reasoning approach on diagnostic accuracy and the consistency of utilization across CPs; and to examine the effect of the diseases; typicality; CP; and problem solving approach on diagnostic accuracy. This study showed similar diagnostic performance between cases that present as chest discomfort and dyspnea (i.e., two similar CPs).

5.1 Discussion related to Research Question 1

What is the underlying structure within clinical cases, and across two similar clinical presentations?

Prior research supported the notion that clinical problem solving is knowledge dependent and case specific; therefore, it cannot be generalized to other cases (Norman et al., 2006; Papa et al., 2007). In other words, diagnostic performance of an expert in chest discomfort cannot predict performance to diagnose a patient complaining of back pain. Research question one examined the degree of case specificity phenomena with clinical cases that present in the same CP and across two similar CPs.

Medical problem solving research led by McGuire and Babbot in the 1960s, assumed that problem solving is a generalizable skill set that can be taught to medical students to apply across different medical problems (McGuire & Babbot, 1967). The concept of generalizable problem solving skills influenced assessment methods until the late 1970s when inconsistency in doctors' performance across clinical cases were observed (Elstein et al., 1978).

Since then, many researchers continuously debated case specificity as an "all or none" phenomena. The findings did not reveal a clear cut line to define the role or quantify the amount of generalizable problem solving skills from one case to another (Eva et al., 1998; Normal et al., 1985; Peverly, 1991).

Norman et al. (2006) published a study titled "How specific is case specificity?" The authors showed that item variance within cases contributed more to the overall variability of a written clinical vignette test than did variance across cases. The recommendation made to enhance test reliability was to increase the number of cases per disease to two or three cases rather than increasing the number of diseases. Although not stated by the authors, these findings challenge the case specificity phenomena.

A study by Mattick et al. (2008) attempted to determine the difference between case-specific and generalizable skills. They estimated that 43-54% of variance accounts for content while 13-16% of variance accounts for generalizable skills. The authors found that history taking and physical examination skill scores were highly correlated within problems and poorly correlated across problems. Norman (2008) argues that the correlations found in this study do not quantify to generalize skills as correlations across cases usually range from 0.10-0.30 and are never 0.00, plus only 3 out the 16 correlations were significant.

In the current study, diagnostic performance scores significantly correlated within and across two similar CP cases. The correlations across cases within a CP ranged from +0.17 to +0.57, indicating a moderate significant linear relationship. Furthermore, the correlation across two similar CPs ranged from +0.17 to +0.66, with 163 statistically significant correlations out of 171. This denotes a significant degree of diagnostic performance generalizability in comparison to the Mattick et. al. (2008) study.

To understand the relationship among cases within and between two CPs, PCA was used as a variable reduction method. Within each CP, diagnostic performance scores for nine to ten cases were assessed. These variables correlated with one another, possibly because they are measuring the same construct. Scores on two similar CPs yielded one PCA component indicating commonality in their diagnostic performance.

Across two similar CPs, low to moderate correlations between cases were observed. A forced two-factor PCA was conducted. The loading of 14 out of 19 diseases on component one resulted in a new dimension similar to the original two CPs. These 14 diseases were a mix of from both CPs and from varying levels of the scheme. Referring to the original CPs, this suggests that most of the diseases did not have a distinctive association with the one CP. Rather, they may be associated with other factors, such as a more general construct of diagnostic accuracy, a compound anatomical or pathophysiologic construct, a technologically-based construct of relatively defining laboratory findings.

Reece et al. (2008) studied OSCE performance scores. Using confirmatory FA, the authors identified best-fit indices with a 3-component model. The authors argue that these three components are history taking, physical examination, and communication skills. Perhaps skills tend to distinct attributes. However, it is possible that diagnostic accuracy may be dependent on several inter-related variables such as disease difficulty, similarity of CPs, and diagnostic strategy used. It would appear that the concept of one general diagnostic strategy was well put to rest.

Wimmers and Fung (2008) used structural equation modeling to characterize OSCE performance assessment score variance. They assumed generalizable skills are imbedded in examinees' traits. Cases as a component of the methods were used to assess the traits. The results

estimated that generalizable skills and case specific knowledge each accounted for 20% of the variance in performance scores. This study found that case difficulty on two similar CPs accounted for the largest amount of variance (29.71%). This amount is greater than that reported by Wimmers and Fung.

Unlike previous studies that reported generalizability of skills from OSCEs, this study focused only on diagnostic accuracy based on written patient scenarios. Previous OSCE studies reported the performance based on history taking; physical examination; and communication skills. These components lack a clear operational definition as they could vary from one OSCE exam to another. For example, physical examination skills can mean observation, palpation, or auscultation.

The performance scores from these studies were given for steps conducted throughout the OSCE. Therefore, an examinee could diagnose a patient incorrectly and still gain scores for the physical examination procedure. In clinical practice, proper diagnosis is the ultimate goal of any medical encounter, which was the main focus within this study.

A second variable that was investigated in this study was problem-solving strategy. For each CP, clerks documented their thinking process while answering the four think aloud exercises provided in Appendix D. Seventy two percent of the clerks in this study used HD reasoning to diagnose the think aloud clinical cases.

As shown in Figure 2, when the knowledge of the diseases is increased but the manifestations of each disease is less developed, the diagnostician is in a dispersed knowledge stage and ultimately uses HD reasoning; that is, testing of one hypothesis at a time for clinical problem solving.

5.2 Discussion related to Research Question 2

How consistent are diagnostic strategies across two cases in the same CP and across two similar CPs?

Past research has indicated that the ability to diagnose a clinical case is enhanced by the clinical problem solving strategy utilized. The root of investigating the expert structure theory originated from studying why chess masters always defeated weekend players (Ericsson et al., 2006). De Groot (1965) identified one factor only, and that is the mental storage of innumerable configurations and the best moves associated with each. He revealed a board configuration from a real game to both chess masters and weekend players and removed it after five seconds. Masters were able to reproduce the board configuration with an accuracy of 70% compared to 30% for weekend players. It takes few minutes to learn chess rules, yet around ten years to master the game by mentally storing board configurations and associating them with the best moves. The mental structures reflected in practice by the moves a player makes on the chessboard.

This same principle may apply to medicine with regards to knowledge structure and problem solving strategy. Diagnosticians that construct complex integrated information together into schemas utilize a SI strategy. In contrast, those who rely on random, isolated facts are more prone to guess or utilize HD reasoning.

The rationale used in many disciplines for teaching problem solving skills is that students can be trained to use general strategies that will help them solve problems in a variety of contexts (Bransford, Sherwood, Vye, & Rieser, 1986; Cramond, Martin, & Shaw, 1990). While the transfer of problem solving skills is tightly linked to knowledge content, Kalyuga and Hanham

(2011) showed evidence of problem solving skill transfer from one context to another with the use of a top-down, general-to-specific approach. In their study, participants had minimal prior knowledge of tasks to be performed, yet still used a general-to-specific approach.

In the current study, the populated sample is from four medical schools that teach and encourage HD reasoning. This problem solving strategy was observed to be the dominant strategy used by 72% of the study population. The transfer of the problem solving skill is evident from classroom teaching context to the test administered. Another important finding from this study is that participants in their last year of medical school showed consistency in use of the same problem solving strategy within cases and across two similar CPs.

This generalizability of diagnostic performance across CPs could be due the fact that clerks` knowledge in chest discomfort and dyspnea are at an advanced level (Papa et al., 2007). Eva et al. (1998) demotes the premise that “content” is the deciding factor of performance generalizability and introduced the importance of using a problem solving strategy. To promote problem solving transfer, the authors suggest: 1) highlighting similarity of diseases to learners; 2) challenging students to engage in higher order, analytical thinking; 3) providing constant guidance and feedback; 4) using the problem solving strategy (e.g., scheme induction) as a learning tool not an algorithm to be blindly followed; and 5) exposing students to cases that link concepts together (Eva, Neville, & Norman, 1998).

Nonetheless, this study provides empirical evidence that students use similar diagnostic strategies across cases from two similar CPs. Further, the findings also indicate that there is no "hop skipping" from a lower to a higher (guessing to scheme induction or vice-versa) diagnostic strategy.

This study supports the findings of similar studies of problem solving strategies and diagnostic performance (Coderre et al., 2003; McLaughlin et al., 2007). ANOVA indicated that participants using a SI problem-solving strategy showed evidence of better diagnostic performance. Thus, students who used scheme induction on average scored 44.7% higher than students who used guessing; 15.5% higher than those who used hypothetical deduction. More importantly, is that these students were all taught HD reasoning throughout their 6 years of medical school. In other words, advanced students will eventually gravitate to SI reasoning in spite of the predominant teaching of HD reasoning. It appears to be a natural outcome.

.

5.3 Discussion related to Research Question 3

What is the effect of disease difficulty, case typicality, CP, and clinical problem solving strategy on diagnostic performance?

Rasch modelling was used to investigate the effect of the above mentioned facets on diagnostic performance. Many studies investigating diagnostic performance used generalizability theory (G-Theory) as a basis for analysis (Mattick, Dennis, Bradley, & Bligh, 2008; Norman et al., 2006; Reece et al., 2008; Wimmers & Fung, 2008). The rationale for using IRT is this study and not G-Theory is based on the following differences between these two measurement frameworks.

The aim of IRT-based Facets and G-Theory is different. As described by Linacre (2013), Facets analyses adopt the philosophy that "raters are independent experts" (p.274); therefore, experts are expected to agree on their overall verdicts, but possibly disagree on details. In contrast, G-Theory`s philosophy is "raters are rating machines", so they are expected to function in the exact same manner. Shavelson and Webb (1991) described the observed scores in G-Theory analyses to be influenced by the "luck of the draw. In Facets, the aim is to provide estimated measures that become the "true" measure after they are adjusted for unwanted source of variance.

In the current study, facets detected elements that were not sufficiently crossed. Data crossing is when facets are not related and do not occur as a smaller model in a larger one, and thus, cannot produce a main effect in the study model. This means that these variables are nested within others.

The findings of this study found typicality and CPs to be nested in diseases. This is similar to sampling leaves from different tree species. The leaves in the current study are diseases. A nested factor would be including another facet identified as the area which the leaf was pulled off- that is, bottom; middle; or top of plant. Location of the leaf sample in this example is similar to typicality levels (prototypical to atypical) and CPs in the current study. In contrast, G-Theory accepts nesting and complex interactions in the study model, which may make the findings difficult to interpret. In MFRM, interactions are deemed to be parted of the error variance and facets are interpreted as main effects.

In G-Theory, item difficulty and item discrimination are dependent on the examinee sample from which they are attained. An above average ability sample will result in higher values of item difficulty. With a below average ability sample lower values of item difficulty will be produced (Hambleton & Swaminathan, 1984). As for item discrimination, heterogeneous examinee samples yield higher values of discrimination, while homogenous examinee sample produce lower values of discrimination. These fluctuations of outcomes by ability of sample are not inherent within IRT. The MFRM facet estimates are independent of the ability of the sample.

Thus, the scores obtained by G-Theory analysis are test dependent. Test difficulty directly affects test scores. Due to this test-dependent property of scores, the same set of test items has to be given to examinees to allow proper comparisons (Hambleton & van der Linden, 1982). It is not possible to use different versions of a test without confounding differences between test scores by differences between the properties of the tests. The best way to adjust for test-dependent scores is through statistical adjustment. Models that statistically adjust for test-dependent scores have been developed in IRT (Prieler, 2007).

In IRT when the assumptions of the model are met, the inherent property of invariance yields item parameter estimates that are independent of the sample of examinees that took the test. Also, examinees' ability estimates are not dependent on the test items answered. The two properties enable comparison of performance on different test forms, provided both groups answer a common subset of questions that measure the same construct (De Champlain, 2010; Crocker & Algina, 2006).

Raw scores of ability and difficulty are usually ordinal in nature. In G-Theory, these raw scores are treated as measurement scales. Meaningful scores are attained by comparison of position in a score distribution (Embretson, 1996; Bond & Fox, 2001).

Interval scale properties are attained by IRT measurement models. Meaningful scores are reached by comparing distances from various items. Ability level and item difficulty are converted and aligned onto an interval scale, which provides a sound way of interpreting the size of gaps between scores (Embretson, 1996; Bond & Fox, 2001). This enables conclusions that person A shows more ability than person B on a certain test by a quantifiable amount. As shown in the current study, candidate 19 had the highest raw score of 69 out of 76 and Rasch true measure of +170, which was not the highest true measure of ability.

In this study, the main concept explored was generalizability of diagnostic performance between two similar CPs. Item discrimination indices from MFRM indicate the extent to which success on an item corresponds to success on the whole test (Linacre, 2013). Any item with negative or zero discrimination undermines the test. Productive items are expected to be positive and closer to one. Very high values indicate that the item repeats information provided by other items on the test.

As shown in Appendix G, sixteen out of nineteen items in this study have estimated discrimination measures above 0.80. These discrimination indices support the findings that performance in a case within the two CPs can help predict performance in another.

5.4 Contributions to Medical Education Literature

This study debates the "all or none" notion of case specificity. The findings of this study indicate that a degree of diagnostic performance predictability exists in the context of cases that are conceptually linked together in a CP and across two similar CPs.

5.4.1 Implications for Testing Agencies

Testing agencies can benefit from the findings of this research in reducing the number of items needed to evaluate competency. In assessment, greater generalizability means fewer items required to test competency. In our current study we found a high degree of predictability of diagnostic performance between cases within and across two similar CPs. Using the Spearman-Brown prophecy formula, which is a formula used to predict the reliability of a test after changing the test length, the items can be reduced. For example, candidate reliability is 0.86 with 19 diseases. Spearman-Brown tells us that a reliability of 0.80 can be achieved with 13 diseases. The diseases to be removed are those with the poorest fit in the Rasch analyses: 6, 10, 17, 3, 16, and 1. However, after dropping those diseases, the reliability is 0.85. Spearman-Brown indicates that 10 diseases are needed to achieve reliability of 0.80. Dropping the three worst fitting diseases (4, 14, and 12) changes the reliability to 0.82. Then Spearman-Brown is re-run, yielding eight diseases. By removing diseases 13 and 19. The reliability becomes 0.80 with eight diseases. Given uni-dimensionality between diseases within the CPs, it is theoretically possible to reduce the number of diseases from 19 to 8 with an expected reliability of 0.80 between the longer and shorten version of the test.

Reducing diseases according to the Spearman Brown formula requires the subset of diseases to have the same distribution of case difficulty as the full set of diseases. The biggest threat is to content validity. The content of the subset of diseases must match the content of the full set of diseases.

Prior research suggested 40 vignettes were needed to obtain reliable assessment of a construct of chest pain (Papa et al., 1998). In this study, as shown is table 11, results showed that it would take around 20-40 cases within a CP to reliably estimate diagnostic performance. If there is a limited amount of testing time, the measure is uni-dimensional, diagnostic performance is analyzed using MFRM, and there is a reduced need to ensure a representative sampling of items within a CP given the generalizability of performance from one case to another, the reduction in cases may be possible This proposition is theoretical and would require further investigation.

Table 11: Number of items required for each CP to obtain an estimated 0.80 reliability.

Study	CP	# of Items	Reliability	Target Reliability	Spearman Brown # of Items
Case et al., 1988	Chest Pain	-	-	0.81	39*
Papa et al., 1998	Chest Pain	-	-	0.80	40*
Current Finding CTT	Chest Pain	36	0.88	0.80	20
Current Finding CTT	Dyspnea	40	0.84	0.80	30
Current Finding MFRM	Chest Pain	36	0.81	-	36*
Current Finding MFRM	Dyspnea	40	0.83	0.80	33

*Not computed by Spearman-Brown prediction formula.

5.4.2 Implications for Medical Schools

The research findings suggest that we can teach physicians to be effective problem solvers only by providing them with domain specific schemas. There is not a general clinical problem solving strategy that can be learned. Clinical reasoning approach is acquired through a large number of specific case problem solving strategies relevant to a particular CP. This recommendation is supported by a randomised trial that supported schema-based instruction as it was associated with better knowledge structure retention and diagnostic performance than HD teaching (Blisset, Cavalcanti, & Sibbald, 2012).

In the current study, the raw mean scores for the diagnostic strategies were: guessing 1.43/4 (35.8%), HD 2.60/4 (65.0%), and SI 3.22/4 (80.5%). Thus, students who used SI on average were 44.7% higher than students who used guessing; 15.5% higher than those who used HD. Blissett et al. (2012) found students who were taught HD or SI generated even a greater difference in performance (38% taught lesions and 31% on untaught lesions). The difference in performance appears to be almost 2 folds in favor of students taught forward reasoning compared to students who are taught backward reasoning. The 2X difference is likely due to differences within the learning environments. Students who used SI in this study were taught within a HD environment while subjects in the Blissett study were directly taught to use SI. Thus, teaching SI can increase the differences in mean outcomes.

5.4.3 Implications for Research

Findings from this study raise a number of questions for future research relative to case specificity and generalizability of diagnostic performance. Research results found high correlations between diagnostic performance within and across cases in two similar CPs. It would be worth examining diagnostic performance predictability using more than two similar and dissimilar CPs.

Typical versus atypical presentation or “textbook presentation” of cases was one of the intervening variables that was anticipated to affect diagnostic accuracy in this research model. This study`s research design failed to capture this variable as a separate entity that is not nested within another factor. Future research can utilize procedures similar to research conducted by Papa and colleagues (1996) to define typicality levels to cases then conduct MFRM to estimate the variance explained by the case typicality factor (Papa et al., 1996).

5.5 Limitations of the Study

This study attempts to examine the structure and variables that affect diagnostic performance. The main limitation is the reliance on pencil and paper scores to estimate diagnostic performance. In reality, clinical performance can include case urgency, chronic illnesses, context, personal attitudes and values, and presence or absence of supporting staff and facilities. A second limitation is the inherent bias in convenience sampling. The population is deemed to be the students of all medical schools in Saudi Arabia. This means that the sample is unlikely to be representative of the population being studied and it undermines the ability to make generalisations from the sample to the population. The sample included participants from four of the 20 Saudi universities that provide an undergraduate medical program. For generalization to be made the four medical schools should have been randomly sampled. Although, these schools appear to representative of all medical schools, this cannot be assumed. The third and last limitation outlined here is the inability to compare the relationship between methods of reasoning instruction (i.e. HD vs. SI) and diagnostic performance. Medical schools involved in the current study all based their learning on hypothetical-deductive instruction. Future studies should compare different schools with different approaches.

5.6 Conclusion

This study answered three research questions concerning diagnostic generalizability and intervening variables that contribute to diagnostic accuracy. Most of the diseases chosen in this study showed high correlation coefficients and high discriminating measures. These estimated measures from two different measurement theories support predictability of performance from one case to another within two similar CPs.

The research model illuminates some factors that influence diagnostic performance. Disease difficulty and diagnostic strategy were two variables found to impact diagnostic performance. Further research is required to identify other influencing variables.

This study also provides evidence that the direction of clinical reasoning utilized enhances diagnostic performance. Schema-based instruction is a recommendation for medical schools to apply as the empirical evidence provided in this study linked better diagnostic accuracy with forward reasoning.

Finally, study findings showed consistent use of the same problem solving strategy within and across two similar CP cases. Generalizability of diagnostic performance and problem solving strategy within and across cases from two CPs are main contributions to the medical education literature that provide ample opportunities for further research.

References

Baghaei, P. (2008). Local dependancy and rasch measures. *Rasch Measurement Transactions, 21,* 1105-1106.

Barrows, H. S. & Bennett, K. (1972). The diagnostic (problem solving) skill of the neurologist: Experimental studies and their implications for eurological training. *Archives of Neurology, 26,* 273-277.

Bechger, T., Maris, G., Verstralen, H., & Beguin, A. (2003). Using Classical Test Theory in combination with Item Response Theory. *Applied Psychological Measurement, 27,* 319-334.

Blisset, S., Cavalcanti, R. B., & Sibbald, M. (2012). Should we teach using schema? Evidence from a randomsied trial. *Medical Education, 46,* 815-822.

Bond, T. (1994). Too many factors? *Rasch Measurement Transactions, 8,* 347.

Bond, T. & Fox, C. (2001). *Applying the rasch model: fundamental measurement in the human sciences*. Psychology Press.

Bordage, G. & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners. *Medical Education, 18,* 406-416.

Bordage, G. (1987). The curriculum: Overloaded and too general? *Medical Education, 21,* 183-188.

Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine, 69.*

Bordage, G. & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Medical Education, 18,* 406-416.

Bordage, G. (2007). Prototypes and semantic qualifiers: From past to present. *Medical Education, 41,* 1117-1121.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist, 41,* 1078-1089.

Case S.M. & Swanson, D. B. (2001). Constructing written test questions for the basic and clinical sciences. *Philadelphia, National Board of Medical Examiners*.

Case, S. M., Swanson, D. B., & Becker, D. F. (1996). Verbosity, window dressing, and red herrings: Do they make a better test item? *Academic Medicine, 71*.

Case, S. M., Swanson, D. B., & Stillman, P. L. (1988). Evaluating diagnostic pattern recognition: the psychometric characteristics of a new item format. *Research in medical education : proceedings of the annual Conference on Research in Medical Education, 27,* 3-8.

Case, S. M. & Swanson, D. B. (1993). Extended-matching items: A practical alternative to free-response questions. *Teaching and Learning in Medicine: An International Journal, 5,* 107-115.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K.A.Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY, US: Cambridge University Press.

Claessen, H. F. A. & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education, 19,* 61-67.

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

Coyle, D. (2009). *The talent code: greatness isn't born. it's grown. here's how*. Bantam.

Cramond, B., Martin, C. E., & Shaw, E. L. (1990). Generalizability of creative problem solving procedures to real-life problems. *Journal for the Education of the Gifted, 13,* 141-155.

Crocker, L. & Algina, J. (2006). *Introduction to classical and modern test theory*. Wadsworth Pub Co.

Daley, B. J. (1999). Novice to expert: An exploration of how professionals learn. *Adult Education Quarterly, 49,* 133-147.

De Champlain, A. (2010). A primer on Classical Test Theory and Item Response Theory for assessments in medical education. *Medical Education, 44,* 109-117.

Dillon, G. F., Boulet, J. R., Hawkins, R. E., & Swanson, D. B. (2004). Simulations in the United States Medical Licensing Examination (USMLE). *Quality and Safety in Health Care, 13,* 41-45.

Downing, S. M. (2003). Item Response Theory: Applications of Modern Test Theory in medical education. *Medical Education, 37,* 739-745.

Elstein, A. & Schwarz, A. (2002). Evidence base of clinical diagnosis: Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ, 324,* 729-732.

Elstein, A., Shulman, L., & Sprafka, S. (1978). *Medical problem solving: an analysis of clinical reasoning*. Harvard University Press.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving. *Evaluation & the Health Professions, 13,* 5-36.

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8,* 341-349.

Ericsson, A., Krampe, R., & Tesch-R, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance (Cambridge Handbooks in Psychology)*. Cambridge University Press.

Ericsson, K. A. & Lehmann, A. C. (1996). Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints. *Annual Review of Psychology, 47,* 273-305.

Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical Education, 41,* 1124-1130.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215-251.

Eva, K. W., Neville, A. J., & Norman, G. R. (1998). Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Academic Medicine, 73*.

Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58,* 357-381.

Field, A. (2000). *Discovering statistics using SPSS for Windows*. Sage Publications.

Fowell, S. L., Maudsley, G., Maguire, P., Leinster, S. J., & Bligh, J. (2000). Student assessment in undergraduate medical education in the United Kingdom, 1998. *Medical Education, 34,* 1-49.

Fox, L. J., Phung, P. C., & Waycott, J. (2006). Effects of advance organizers, mental models, and abilities on task and recall performance using a mobile phone network. *Applied Cognitive Psychology, 20,* 1143-1165.

Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology, 4,* 261-272.

Gladwell, M. (2008). *Outliers: the story of success*. Little, Brown and Company.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Grant, J. & Marsden, P. (1987). The structure of memorized knowledge in students and clinicians: An explanation for diagnostic expertise. *Medical Education, 21,* 92-98.

Hambleton, R. & Swaminathan, H. (1984). *Item Response Theory: principles and applications (evaluation in education and human services)*. Springer.

Hambleton, R. K. & Jones, R. W. (1993). An NCME instructional module. *Educational Measurement: Issues and Practice, 12,* 38-47.

Hambleton, R. K. & van der Linden, W. J. (1982). Advances in Item Response Theory and applications: An introduction. *Applied Psychological Measurement, 6,* 373-378.

Harasym, P. H., Norris, D. A., & Lorscheider, F. L. (1980). Evaluating student multiple-choice responses: Effects of coded and free formats. *Evaluation & the Health Professions, 3,* 63-84.

Harasym, P. H., Tsai, T. C., & Hemmati, P. (2008). Current trends in developing medical students' critical thinking abilities. *The Kaohsiung journal of medical sciences, 24,* 341-355.

Hardie, E. M. (2008). Current methods in use for assessing clinical competencies: What works? *Journal of Veterinary Medical Education, 35,* 359-368.

Hobu, P. P. M., Schmidt, H. G., Boshhuizen, A., & Patel, V. L. (2009). Contextual factors in the activation of first diagnostic hypotheses: expert novice difference. *Medical Education, 21,* 471-476.

Hoffman, R. R. (1988). How can expertise be defined?: implications of research from cognitive psychology. In Williams R, Faulkner W, & Fleck J (Eds.), *Exploring Expertise* (pp. 81-100). Edinburgh, Scotland: University of Edinburgh Press.

Kalyuga, S. & Hanham, J. (2011). Instructing in generalized knowledge structures to develop flexible problem solving skills. *Computers in Human Behavior, 27,* 63-68.

Kim, S. I., Swanson, T. A., & Caplan, J. P. (2007). *Underground clinical vignettes step 2*. (Fourth ed.) Lippincott Williams and Wilkins.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. D. (2000). *To err is human: building a safer health system*. Washington, D.C.: National Academu Press.

Kushniruk, A. W., Patel, V. L., & Marley, A. A. (1998). Small worlds and medical expertise: implications for medical cognition and knowledge engineering. *Int J Med Inform, 49,* 255-271.

Landon, B. E., Reschovsky, J., Reed, M., & Blumenthal, D. (2001). Personal, organizational, and market level influences on physicians' practice patterns: Results of a national survey of primary care physicians. *Medical Care, 39*.

Leape, L. L. (1994). Error in medicine. *JAMA: The Journal of the American Medical Association, 21,* 1851-1857.

Linacre, J. (1998). Rasch first or factor first? *Rasch Measurement Transactions, 11*, 603.

Linacre, J. (2005). Rasch dichotomous model vs. one-parameter logistic model. *Rasch Measurement Transactions, 19,* 1032.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7,* 328.

Linacre, J. M. (2013). *A user`s guide to FACETS*. Winsteps.

Lord, F. M. (1980). *Applications of Item Response Theory to practical testing problems*. Routledge.

Macdonald, P. & Paunonen, S. V. (2002). A monte carlo comparison of item and person statistics based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement, 62,* 921-943.

Mandin, H., Harasym, P., Eagle, C., & Watanabe, M. (1995). Developing a "clinical presentation" curriculum at the University of Calgary. *Academic Medicine, 70.*

Mandin, H., Harasym, P., & Woloschuk, W. (2000). Clinical problem solving and the clinical presentation curriculum. *Academic Medicine, 75,* 1043.

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

Mattick, K., Dennis, I., Bradley, P., & Bligh, J. (2008). Content specificity: is it the full story? Statistical modelling of a clinical skills examination. *Medical Education, 42,* 589-599.

McGuire, C. H. & Babbot, D. (1967). Simulation technique in the measurment of problem solving skills. *Journal of Educational Measurment,* 1-10.

McLaughlin, K., Coderre, S., Mortis, G., & Mandin, H. (2007). Expert-type knowledge structure in medical students is associated with increased odds of diagnostic success. *Teach Learn.Med, 19,* 35-41.

McLaughlin, K. & Mandin, H. (2002). Using "concept sorting" to study learning processes and outcomes. *Academic Medicine, 77.*

Morris, G., Martin, L., Harshman, N., Baker, S., Mazur, E., Dutta, S. (2006). Testing the test: item response curves and test quality. *American Journal of Physics, 74,* 449-453.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem solving by medical students: A cross sectional and logitudinal analysis. *Medical Education, 15,* 315-322.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical Education, 15,* 315-322.

Norcini, John (2004, November 16). Back to the future: Clinical vignettes and the measurement of physician performance. *Annals of Internal Medicine*.

Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education, 39,* 418-427.

Norman, G., Bordage, G., Page, G., & Keane, D. (2006). How specific is case specificity? *Medical Education, 40,* 618-623.

Norman, G. R. (2008). The glass is a little full of something: revisiting the issue of content specificity of problem solving. *Medical Education, 42,* 549-551.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1166-1174.

Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition and Instruction, 17,* 433-448.

Norman, G. R. & Eva, K. W. (2003). Doggie diagnosis, diagnostic success and diagnostic reasoning strategies: An alternative view. *Medical Education, 37,* 676-677.

Olson, J. R. & Rueter, H. H. (1987). Extracting expertise from experts: Methods for knowledge acquisition. *Expert Systems, 4,* 152-168.

Papa, F. J. & Elieson, B. (1993). Diagnostic accuracy as a function of case prototypicality. *Academic Medicine, 68*.

Papa, F. J. & Harasym, P. H. (1999). Medical curriculum reform in North America, 1765 to the present: a cognitive science perspective. *Academic Medicine, 74,* 154.

Papa, F. J., Harasym, P. H., Schumacker, R., Aldrich, D. G., & Stone, R. C. (1998). The impact of an alternative approach to construct defintion upon the reliability and utility of diagnostic performance measures. *Academic Medicine, 73,* 100-102.

Papa, F. J., Stone, R. C., & Aldrich, D. G. (1996). Further evidence of the relationship between case typicality and diagnostic performance: implications for medical education. *Academic Medicine, 71*.

Papa, F., Oglesby, M., Aldrich, D., Schaller, F., & Cipher, D. (2007). Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. *Medical Education, 41,* 419-425.

Patel, V. L. & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10,* 91-116.

Patel, V. L., Groen, G. J., & Frederiksen, C. H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education, 20,* 3-9.

Patel, V. L., Groen, G. J., & Scott, H. S. (1988). Biomedical knowledge is explanations of clinical problems by medical students. *Medical Examiner, 22,* 398-406.

Peabody, J. W. & Liu, A. (2007). A cross-national comparison of the quality of clinical care using vignettes. *Health Policy Plan, 22,* 294-302.

Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA: The Journal of the American Medical Association, 283,* 1715-1722.

Peabody, J. W., Luck, J., Glassman, P., Jain, S., Hansen, J., Spell, M. (2004). Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Annals of Internal Medicine, 141,* 771-144.

Peabody, J. W., Tozija, F., Munoz, J. A., Nordyke, R. J., & Luck, J. (2004). Using vignettes to compare the quality of clinical care variation in economically divergent countries. *Health Services Research, 39,* 1951-1970.

Peckler, B., Schocken, D., & Paula, R. (2009). Simulation in a high stakes clinical performance exam. *Journal of Emergencies, Trauma and Shock, 2,* 85.

Prieler, J. (2007). So wrong for so long: Changing our approach to change. *The Psychologist, 20*.

Reece, A., Chung, E. M. K., Gardiner, M. R., & Williams, S. E. (2008). Competency domains in an undergraduate Objective Structured Clinical Examination: their impact on compensatory standard setting. *Medical Education, 42,* 600-606.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing, 25,* 465-493.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: theory and implications. *Acad Med, 65,* 611-621.

Schmidt, H. & Boshuizen, H. (1993). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21,* 338-351.

Schmidt, H. G. & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education, 41,* 1133-1139.

Schuwirth, L. W. T., Van Der Vleuten, C. P. M., Stoffers, H. E. J. H., & Peperkamp, A. G. W. (1996). Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Medical Education, 30,* 50-55.

Schuwirth, L. W. T. & van der Vleuten, C. P. M. (2004). Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education, 38,* 974-979.

Shavelson, R. J. & Webb, N. M. (1991). *Generalizability Theory: a primer*. Thousand Oaks: SAGE.

Tekian, A., McGuire, C. H., & McGaghie, W. C. (1999). *Innovative simulations for assessing professional competence: from paper and pencil to virtual reality*. Chicago: Univ of Illinous at Chicago Dept.

Tennant A. & Pallant J.F. (2006). Unidimensionality matter! (a tale of two smiths?). *Rasch Measurement Transactions, 20,* 1048-1051.

Tsuen-Chiuan, T., Harasym, P. H., Coderre, S., McLaughlin, K., & Donnon, T. (2009). Assessing ethical problem solving by reasoning rather than decision making. *Medical Education, 43,* 1188-1197.

Veloski, J. J., Rabinowitz, H. K., & Robeson, M. R. (1993). A solution to the cueing effects of multiple choice questions: the Un-Q format. *Medical Education, 27,* 371-375.

Veloski, J., Tai, S., Evans, A. S., & Nash, D. B. (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. *American Journal of Medical Quality, 20,* 151-157.

Wimmers, P. & Fung, C. (2008). The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Medical Education, 42,* 580-588.

Woloschuk, W., Harasym, P., Mandin, H., & Jones, A. (2000). Use of scheme-based problem solving: an evaluation of the implementation and utilization of schemes in a clinical presentation curriculum. *Medical Education, 34,* 437-442.

Wright, B. & Linacre, J. (1989). Differences between scores and measures. *Rasch Measurement Transactions, 3,* 63.

Wright, B. D. (1996). Local dependancy, correlations, and principal components. *Rasch Measurement Transactions, 10,* 509-511.

Appendices

Appendix A: Review of theory types used to explain expertise in medicine

Author(s)	Type of Theory	Main Findings
Barrows & Bennet, 1972	Process	Problem solving was characterized by early hypothesis generation and hypothetico-deductive reasoning
Elstein et al. 1978	Process	Both experts and novices used hypothetico-deductive reasoning for problem solving
Neufeld et al. 1981	Process	Hypothetico-deductive reasoning used by all subjects at different expertise levels; Experts generated better early hypothesis that predicted final diagnosis more accurately
Patel & Groen, 1986	Process / Memory	Experts use forward clinical reasoning / Unsuccessful in analyzing data base and recall
Claessen & Boshuizen, 1985	Memory	No superior memory performance between experts and novices
Patel et al. 1986	Memory	Experts are superior in recalling relevant clinical information
Norman et al. 1989	Memory	Intentional memory is a misleading measure of expertise
Schmidt & Boshuizen, 1993	Memory	Experts have access to extensive case knowledge that is encapsulated until needed
Bordage & Zacks, 1984	Structure (Prototype)	Prototype categorization of knowledge in memory
Grant & Marsden, 1987	Structure	Experts is based on individualized memory structures
Schmidt et al. 1990	Structure	Knowledge structure evolves as: elaborated causal networks, abridged causal networks, Illness scripts, Instance scripts.
Bordage, 1994	Structure (semantic qualifiers)	Knowledge structure evolves as: reduced, dispersed, elaborated, and compiled.
McLaughlin & Mandin, 2002	Structure	Use of schemes led to higher concept sorting scores
McLaughlin et al. 2007	Structure	Use of schemes led to increased number of expert type concepts in knowledge structure, which led to better diagnostic accuracy

Appendix B

Appendix B1: Clinical Vignettes

INSTRUCTIONS:

Read each clinical vignette carefully. Each patient scenario describes one disease. From the list of choices provided select the disease that best describes the patient's presenting signs, symptoms, and laboratory information. There is only one correct answer for each scenario.

Indicate your answer by filling the circle that corresponds to the letter (A-S) of your answer on the a

1- A 27 year old man presents with chest pain and low grade fever of 1 week`s duration. His pain is described as retrosternal and radiates to his back, worsens with deep inspiration, and is relieved when he sits up and leans forward. Vitals: normal. Physical Examination: tachycardia with regular rhythm; triphasic, high pitched scratching sound heard over left lower sternal border.
Labs- ECG: PR segment depression and diffuse ST segment elevation with an upright T wave.
Imaging- Echo: good cardiac function with small pericardial effusion. CXR: normal.

2- Dave is a 20 year old man, previously well, presents with sudden onset of pleuritic chest pain and breathlessness. His oxygen saturations are 96% and he has a respiratory rate of 16 breaths/min. His blood pressure and heart rate are normal. He has reduced air entry and hyper-resonance in the right chest. His trachea is central.

3- A 65-year-old woman presents to the ER with chest pain, fever and hypoxia. Yesterday she developed a cough productive of yellow-green sputum and mild fever. She also complains of fatigue, and malaise.
Vitals: fever (38.6 C); hypotension (BP 95/54); tachycardia (HR 101); tachypnea (RR 32); O2 sat of 91% on 3 L O2. Physical Examination: elderly female in severe respiratory distress, using accessory respiratory muscles. Lung exam demonstrates rales and bronchial breath sounds in the right upper lung field, egophony, and dullness to percussion.
CBC: marked leukocytosis, Basic Metabolic Panel: decreased HCO3 (15). ABG: pH 7.22/PCO2 38/PO2 65 (metabolic acidosis with hypoxia). Lactate elevated. Sputum culture: many PMNs, few epithelial cells, and gram-negative rods (identified as Haemophilius influenza after 2 days). Blood cultures: negative.

4- A 58-year-old African American male was seen in the ER for bronchitis a year ago. The CXR showed a right upper lobe mass, and the patient was told to follow-up with his physician but he never did. He now complains of cough productive of yellow sputum for a week and reports some weight loss but he cannot say how much. Smoker with a 30 pack-year history, quit 2 weeks ago. He drinks a pint of liquor per day.
Vitals: T:37.7- HR: 105-BP: 137/61-BR; 18. Chest: breath sounds are reduced in the right upper lobe. CVS: Clear S1S2. Extremities: clubbing.
Labs- CBC shows anemia due to multiple factors: suspected iron deficiency, malnutrition, alcohol abuse and chronic disease.
Imaging: CT Chest: There is a large relatively hypodense mass involving much of the right upper lobe and directly abutting the superior mediastinum on the right.

5- A 10 year old boy presents with difficulty breathing, cough, and wheezing. His symptoms started 4 days ago when he contracted an upper respiratory infection. He is allergic to cats and pollen and had eczema and allergic rhinitis during infancy. His father smokes cigarettes.
Vitals: slight fever; tachypnea; tachycardia. Physical examination: unable to speak in full sentences; nasal mucosa boggy and pale; pulsus paradoxus present; moderate supraclavicular, intercostal, and subcostal

retractions; decreased air movement throughout all lung fields with scattered expiratory wheezes; prolonged expiratory phase; hyperresonant to percussion.
Labs- CBC: leukocytosis with lymphocytosis and eosinophilia. Increased serum IgE. ABG: respiratory alkalosis, hypocarbia, and hypoxemia. PFTs: increased total lung capacity and residual volume; decreased FEV1/FVC.
Imaging- CXR: lung hyperinflation with flattened diaphragms; peri-bronchial cuffing and linear atelectasis in both bases.

6- A 56-year-old man presents with shortness of breath and a chronic cough productive of copious mucoid sputum. Over the past 3 years, his symptoms have occurred for at least 3 months of every year. He also reports a 40-pack-year smoking history.
Vitals: tachypnea (RR 25). Physical examination: blue lips; plethora; bilateral ronchi and wheezes.
Labs- CBC: hematocrit increased. ABGs: hypoxia and hypercapnia. ECG: presence of P pulmonale and poor progression of R wave in chest leads. PFTs: obstructive pattern of increased residual volume and decreased FEV1/FVC. Imaging- CXR, PA: presence of increased basilar bronchovascular markings and thickened bronchial walls.

7- A 56 year old man presents with a 10 month history of increasing dyspnea on exertion and occasional dry cough. He has been smoking half a pack of cigarettes daily for 40 years and has a history of rheumatoid arthritis. Chest examination reveals mild hyper-resonance in all lung fields and diminished breath sounds. Lung volume measurements show increased total lung capacity (TLC) and residual volume (RV), with an elevated RV:TLC ratio.

8- A 74 year old woman had difficulty breathing and suddenly became delirious 1 hour ago. She was well until 5 days ago, when she tripped and injured her right leg; she has been bedridden since the accident. On examination, her temperature is 37.2 degrees Celsius, blood pressure is 110/70 mm Hg, pulse is 110/min, and respirations are 32/min. Her heart, lungs, and abdomen are normal, and there are no focal motor or sensory deficits. Passive flexion of her right hip causes obvious pain. Pulse oximetry shows an oxygen saturation of 80%, and a chest x-ray is normal. An ECG shows sinus tachycardia.

9- A 40 year old woman complains of progressive shortness of breath (SOB) for 2 weeks. She has a history of melanoma and hypertension. Initially her SOB was on exertion, but over the past 2 days it has been present even at rest. She denies fever, chills, orthopnea, or paroxysmal nocturnal dyspnea.
Vitals- tachycardia (HR 130), tachypnea (RR 30). Physical examination: positive jugular venous distension; regular rate and rhythm; distant heart sounds; pulsus paradoxus; bilateral inspiratory crackels and 2+ lower extremity edema.
Labs- CBC: normocytic anemia; lytes: hyponatremia; mild creatinine elevation. Normal cardiac enzymes. Urinalysis: normal
Imaging- ECG: regular rhythm, sinus tachycardia, low voltage QRS, electrical alternans.

10- A 29 year old woman presents to the clinic complaining of shortness of breath, persistent fatigue, and weakness. She states that she has heavy menstrual bleeding and frequent bleeding between periods. She adds that she has brittle nails and frequently eats ice.
Vitals: normal. Physical examination: pallor; no scleral icterus; moist mucous membranes; atrophic tongue; 2-sec capillary refill; regular rate and rhythm.
Labs- CBC: hypochromic, microcytic RBCs; low hemoglobin (9 g/dL) and hematocrit (28%); elevated RDW; low MCV and MCHC. Ferritin low; TIBC elevated.

11- A 2 year old African American boy presents with shortness of breath and painful swelling of the hands and feet. The child has had recurrent abdominal and joint pain. He has a mild fever, headache, shortness of breath, and a feeling of heaviness in his abdomen.
Vitals: fever (38.3 C) Physical examination: in pain; pallor of conjunctiva, skin, and mucous membranes; mild icterus; hypoxic spots with neovascularization on retina; throat hyperemic; heart tachycardic with regular

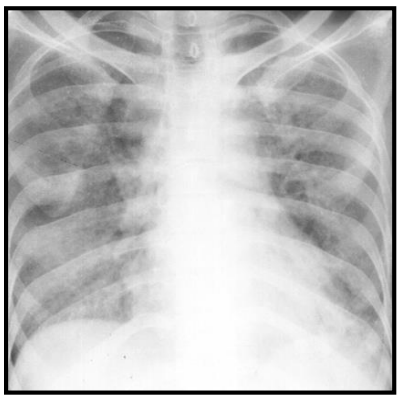
rhythm with 2/6 flow murmur at LSB; lungs clear; splenomegaly present; skin over metacarpals and metatarsals is warm to touch; joints tender.
Labs- CBC: hemoglobin low (7), reticulocytes high (8.3%).
Imaging- XR, hands and feet: soft tissue swelling with radiolucent areas.

12- A 37-year-old man presents complaining of chronic and recurrent upper abdominal pain. The pain is burning in quality, occurs when the stomach is empty, and is relieved within minutes by food or antacids. He doesn't have evidence of gastrointestinal bleeding or anemia. He is taking indomethicin for gout. Medical records show that he had a gout flare up 3 months ago and was discharged home with a steroid taper. He was prescribed Indomethacin 50 mg po q 8 hr PRN pain but he was taking it daily for the last month.

13- Jill is a 50-year-old woman who lives with her husband and two children (aged 20 and 18). She has come to see her GP with worries about a number of chest discomfort, extreme tiredness, and agitation. Jill has been a frequent attender at the practice over the years, often with concerns about her or her children's health. She has been on and off antidepressants for the past 30 years. When she was 23 she took an overdose following the break-up of a relationship. She had some sessions of counselling about 10 years ago that she found helpful. During the past 8 months following her husband's diagnosis of heart problems. She has been drinking wine most evenings to try to calm herself down.
Vitals: normal. Physical examination: blond and fair-skinned; no acute distress.
Labs- CBC/Lytes: normal.
Imaging: ECG, CXR: normal.

14- A 32 year old woman complains of palpitations, diaphoresis, dyspnea and fear when riding elevators. She works for a large firm in the financial district and was recently promoted. Her promotion has taken her from an office on the 9th floor to an office on the 32nd floor of the building. She has always taken the stairs to her office, but now this has made her late for numerous meetings, drawing the attention of her supervisor. She reports normal sleep, appetite, and energy levels. There is no past history of traumatic events. Vitals: stable. Physical examination: normal. Labs- Lytes/CBC: normal. UA: toxicology negative.

15- A 35-year-old patient presents with complaints of high grade fever for 3 days and a cough with mucoid expectoration for 2 days. On examination pulse is 96 bpm and blood pressure is 90/70 mmHg. On general physical examination there is pallor present. Abdominal examination reveals mild splenomegaly only. On investigation the peripheral smear is positive for plasmodium falciparum. The patient is started on IV quinine on the basis of peripheral smear report. However on treatment the patient develops progressive dyspnea. The ABG shows a lower PaO_2 and the chest x-ray:



16- A 61 year old woman presents to her family doctor complaining of progressive shortness of breath and noticeable ankle swelling. She has a known history of hypertension and dyslipidemia; she has chronic stable

angina that is effort dependant and has not had a prior myocardial infarction. She is currently receiving Aspirin 325 mg qd; atorvastatin 40mg qd; amlodipine 10 mg qd; 10 mg qd; and atenolol 50 mg BID.
Vitals- BP 160/105; P 88 regular. Physical examination: BMI 31 kg/m2; bibasilar crackles are noted; normal S1S2; obese abdomen; 1+ pedal edema.
Lab- Elevated BUN and blood creatinine levels; normal potassium. ECG: NSR with non-specific ST/T wave changes.
Imaging- Stress Echo: no angina reported; exercised 5 min of modified Bruce protocol; peak HR 140 bpm; diffuse but minor ST downslopping; baseline LVEF of 0.50; LVH is present with impaired LA emptying; no significant wall motion abnormalities with exercise.

17- A 34 year old white obese female complains of shortness of breath, dizziness, and near fainting spells. She has been taking prescription medication for approximately 6 months in order to lose weight. Upon examination the patient is obese and looks cyanotic; large "a" wave in jugular venous pressure; parasternal heave; loud S2; narrow splitting of S2; rales on both bases; hepatomegaly. Labs: CBC- polycythemia. ECG- right axis deviation; right ventricle and right atrial hypertrophy. ABGs- hypoxemia. Imaging: Chest X-ray- enlarged right ventricle; enlarged main pulmonary artery with peripheral pruning.

18- A 25 –year- old woman suffers from progressive shortness of breath for the past 5 years. On several occasions she brought out blood tinged sputum. She has suffered from nonspecific joint pains. On examination, she is dyspnoeic and mildly cyanosed. There is left parasternal heave and a harsh systolic murmur heard over precordium and her p2 is loud. There is no click opening snap. Her condition has further deteriorated. ECG shows RBBB pattern.

19- A 62-year-old African American male is admitted to the hospital with shortness of breath (SOB) for three days. Past medical history of Congestive heart failure and hypertension.
Vitals: BP 75/40 mmHg. Physical Examination: HEENT: + JVD. Heart: distant S1S2.
Chest: Clear symmetrical breath sounds bilaterally.
Imaging: Chest X-ray: massive bottle-shaped heart and conspicuous absence of pulmonary vascular congestion. Echo: Severe global impairment of left ventricular systolic function with relative preservation of the base of the heart; moderate concentric left ventricular hypertrophy; large pericardial effusion which appears to be hemodynamically compromising the heart.

20- A 71-years-old woman was referred a specialist because of anterior acute myocardial infarction, about 6 hours after the onset of symptoms. Electrocardiogram showed sinus rhythm and ST elevation in precordial V1 → V4 leads. She did not complain dyspnea.
Coronary arteriogram showed a sub-occlusion (TIMI 1–2) of mid-portion of the left anterior descending artery, which was successfully treated with percutaneous coronary intervention with stent implantation.
On auscultation a 3/6 grade systolic murmur was heard at the lower left sternal border.
Echocardiogram evidenced an hypertrophic left ventricle of normal dimension, with apical akinesis, ejection fraction was 48%. Moreover posterior mitral annulus appeared calcified with a rounded, vacuolated mass of 1 cm in diameter (Fig. 2). Transesophageal echocardiogram: mitral posterior leaflet was thickened and partially calcified, otherwise chordate tendineae appeared normal but the posterior valvular ring and the base of leaflet was deformed by a cystic mass consistent with caseous calcification.

21- A 60 year old man complains of shortness of breath with exertion and had one episode of syncope while playing football. The patient states that he has experienced increasing fatigue and a decrease in his normal activity. He adds that he sometimes feels left sided chest pain induced by exercise.
Vitals: tachycardia; narrow pulse pressure. Physical examination: delayed, slow rising carotid upstroke, forceful apical beat; soft S2; harsh, late peaking systolic ejection murmur that is loudest at right second intercostal space, radiating to carotids.
Labs – ECG: left ventricular hypertrophy; ST depression and T-wave inversion.
Imaging – Echo: bicuspid aortic valve; left ventricle wall thickening; decreased valvular area.

22- A 64-year-old man presents to the ER complaining of sudden-onset substernal chest pain that awakened him from sleep, the pain has lasted for at least 30 minutes. He describes his pain as a pressure radiating to the left arm that is accompanied by mild shortness of breath, diaphoresis, and nausea. Nitroglycerin only transiently relieved his symptoms, and postural changes and inspiration did not alter his chest pressure. His past medical history is significant for hypertension and hyperlipidemia and his brother died of an MI at 51 years of age. He additionally reports to a 20-pack-year smoking history. Vitals: hypertension (BP 148/68); tachycardia (HR 101); tachypnea (RR 30). Physical Examination: obese, pale and diaphoretic; SaO2 96% on room air; lungs clear; S4 gallop noted; no cyanosis or edema. CBC: normal hematocrit; mild leukocytosis. Lytes: normal. Elevated troponin T and I; CK-MB levels elevated 4 to 6 hours after pain onset.

23- A 30-year-old patient presents with a 2-month history of fever, exertional dyspnea, and fatigue. The patient denies a history of cough, chest pain, jaundice, or oliguria. On examination, there is mild pallor but no icterus, cyanosis, or clubbing. The patient's jugular venous pressure is raised with Y-descent more prominent than X-descent. Kussmaul's sign is positive, the apical impulse is feeble with retraction noted in systole, and there is no murmur. The patient has ascites and hepatomegaly. There is no splenomegaly. Mild pedal edema is present.

24- Saly presents to her primary care physician with complaints of episodes of chest discomfort. She is a 62-year-old divorced mother of two who has been suffering from increasing chest discomfort for the past 1-2 years. At first she only noticed the discomfort when she experienced extreme exertion, such as running for the bus, and these symptoms always promptly resolved when she stopped or slowed down. She attributed this to no more than "getting old" and "being out of shape." About 6-9 months ago, she noted that the chest pain episodes had become more frequent, occurring perhaps once or twice per week and precipitated by stress, either physical (walking up and down stairs at home with the laundry) or mental (especially after a day's work at her job in a government office), but again always resolved with rest. When the chest discomfort episodes began to occur more than once per week, she was persuaded to visit her primary care physician. She is overweight and smoked cigarettes 1 pack/day for 15 years, although she managed to taper off and completely stopped about 5 years ago. She drinks a "very occasional" glass of wine. She admits to having a "poor diet" consisting of high-salt, high-fat foods, with few fresh vegetables or fruits, but reports that she takes multivitamins to compensate for her poor diet.
She has a history of hypertension (blood pressure, 160/92 mm Hg), which was diagnosed in her late 40s, and she has been relatively adherent to a prescription of enalapril 10 mg and "a diuretic" for the last 10+ years. Dyslipidemia (total cholesterol, 215 mg/dL; LDL, 138 mg/dL; HDL, 35 mg/dL) was diagnosed during a clinic visit 2 years ago; she takes lovastatin 10 mg daily "when she remembers."

25- A 47 year old man reports a six month history of intermittent chest discomfort while playing squash. He describes lower substernal tightness with numbness of the left upper arm only during exertion. He does not smoke. His father died suddenly at the age of 49 years.
Vitals- BP: 138/84 mmHg. Physical examination: normal.
Labs- The level of total cholesterol 261 mg/dL (n: <200); of LDL 172 mg/dL (n:60-130); of HDL 50 mg/dL (n: >40), and the triglyceride level is 113mg/dL (n:10-150).
The result of an exercise test is positive, with pain and 1.5 mm of horizontal ST-segment depression at stage 4 of the Bruce protocol.

26- A 25 year old white male complains of sudden pleuritic chest pain and shortness of breath that awakens him at night. He smokes one pack of cigarettes a day and states that his paternal uncle once had a similar episode. On examination, he was a tall, thin patient; diaphoretic and feels weak; left chest expands weekly on inspiration; trachea and apex beat displaced to right; left side hyper-resonant to percussion; decreased breath sounds; decreased tactile fremitus. ABGs showed decreased PO2 and elevated PCO2. Chest X- ray revealed a partial collapse of left lung with no lung markings except thin line parallel to chest wall; costophrenic sulcus abnormally radiolucent in supine film.

27- A 67-year-old man presents to the clinic reporting intermittent upper chest and jaw pain lasting 2 to 3 minutes for several weeks. Initially, the episodes were infrequent and only occurred with exertion. Recently, the episodes have become more frequent and now are occasionally present at rest. The pain is associated with shortness of breath and diaphoresis. The patient has hypertension and type 2 DM, with a recent HbA1c of 7.9. He has a heavy smoking history, drinks alcohol, and leads a sedentary life. The patient is currently free of chest pain. Vitals: hypertension (BP 160/80); HR 74; RR 18, O2 sat 98% on room air, Physical Examination: obese, well-appearing man with an unremarkable exam, except for decreased pedal pulses bilaterally. CBC and BMP: normal. CK-MB and troponin normal. ECG: normal sinus rhythm with LVH and signs of strain; unchanged from prior ECG. CXR: unremarkable.

28- A 42 year old man business executive complains of recurrent dyspepsia over the past month. The patient reports epigastric discomfort that he describes as gnawing, dull, achy, intermittent, episodic, that is relieved by food or antacids. He acknowledges a high stress lifestyle with frequent business travel; he also smokes two packs of cigarettes per day and drinks heavily.
Vital signs: normal. Physical examination: abdomen soft; mild tenderness to deep palpation in midepigastrium; no occult blood in stool.
Labs- CBC: hematocrit normal. Rapid urease breath test positive; serum Helicobacter pylori antibody positive; gastric mucosal biopsy reveals H. pylori infection; no evidence of malignancy.

29- A 30-year-old patient presents with acute onset of dyspnea. PaO2/FI O2 is 190 mmHg, while chest x-ray shows bilateral alveolar infiltrates. Pulmonary capillary wedge pressure is 16 mmHg.

30- A 22-year-old woman suspected of having primary Sjögren‘s syndrome was referred to the Department of Maxillofacial Surgery complaining of bilateral swelling of the eyelids and the parotid glands, also complained dry mouth and eyes of three weeks duration and visual disturbance that developed in the evenings. She denied fatigue, fever, night sweats or any other sign of general malaise, otherwise, the patient was in good general health.
Physical examination: bilateral enlargement of both parotid glands, which were solid, plain and painful to pressure, especially right side. Three days after admission the patient developed right facial nerve palsy and taste disturbance developed.
Labs- Serologic investigations: S-HZV, S-CMV, Epstein Barr virus (EBV), mumps virus IgM negative.
Imaging- CXR: bihilar lymphadenopathy. Polycystic circumscribed enlargement of pulmonal hiluses. CT: bilateral parotid gland enlargement and destruction foci were observed on the right gland tissue.
Excision biopsy from the small salivary glands of lower lip: Nodules consisting of epithelioid cells, surrounded by dense fibrous tissue and centrally Langerhans´ type giant cell is observed. There is no evidence of caseation.

31- A 39 year old man presents to the emergency department with acute onset of shortness of breath, hemoptysis, and left sided pleuritic chest pain. His past medical history includes medication controlled asthma, peptic ulcer disease, and a recent onset of idiopathic nephrotic syndrome. His blood pressure is 180/100 mm Hg, pulse is 110/min, and respirations are 28/min. Cardiac and lung examinations are normal. Laboratory data are remarkable for a serum lactate dehydrogenase of 300U/L. An ECG shows sinus tachycardia, prominent S waves in lead I, inversions of the T wave, and a prominent Q wave in lead III.

32- A 38 year old black woman presents with the insidious onset of shortness of breath, chest pain, and fatigue. Physical examination reveals enlarged cervical lymph nodes and scattered brown-red papules on the skin. A chest x-ray film shows bilateral pulmonary infiltrates and enlarged hilar lymph nodes. Biopsies of skin lesions and lymph nodes show non-necrotizing granulomas. Special stains for fungi and mycobacteria are negative.

33- A 34-year-old woman complains of difficulty breathing. She has a history of asthma and reports 1 week of productive cough and rhinorrhea. She also describes fevers, chills, and sweats. She had a prior intubation for an asthma exacerbation. Medications include albuterol and flunisolide at home.
Vitals: tachycardia (HR 130), pulsus paradoxus; tachypnea (RR 38); fever (38.2C); SaO2 70% on room air.
Physical Examination: well-developed, well-nourished female with severe respiratory distress and mild cyanosis; no tonsillar adenopathy or erythema; trachea is mid-line; absent lung sounds in bilateral posterior bases; faint expiratory wheezing with increased expiratory-to-inspiratory (E:I) ratio; heart tachycardic with no gallops; accessory muscles used for breathing.
CBC: WBC elevated with left shift; eosinophilia. CXR: right lower lobe infiltrate.

34- A 40 year old woman complains of progressive shortness of breath (SOB) for 2 weeks. She has a history of melanoma (diagnosed 10 years ago, s/p surgical resection) and hypertension. Initially, her SOB was on exertion, but over the past 2 days it has been present even at rest. She denies fever, chills, orthopnea, or paroxysmal nocturnal dyspnea. Vitals: tachycardia (HR 130), tachypnea (RR 30). Physical examination: positive jugular venous distension; regular rate and rhythm; distant heart sounds; pulsus paradoxus; bilateral inspiratory crackles and 2+ lower extremity edema.
Labs – CBC: normocytic anemia; lytes: hyponatremia; mild creatinine elevation. Normal cardiac enzymes. Urinalysis: normal.
Imaging – ECG: regular rhythm, sinus tachycardia, low voltage QRS, electrical alternans.

35- An 82-year-old woman, with history of hypertension, was admitted for paroxysmal atrial fibrillation, which was successfully treated with intravenous Amiodarone. Physical examination revealed a systolic murmur of 2–3/6 L grade. Besides laboratory examinations were substantially normal.
Echocardiogram demonstrated mild left atrial enlargement (diameter = 42 mm, area = 22 cm^2), therefore left ventricle was hypertrophic with an ejection fraction of about 60%. Additionally in the posterior site of mitral annulus, at the base of corresponding leaflet, was a round cystic mass of about 2 cm in diameter, with a calcified capsule and an echo-lucent core. The lesion was consistent with caseous calcification of mitral annulus.
Color-Doppler examination ruled out shunt between the cyst and the cardiac chambers, moreover color flow mapping revealed a moderate mitral insufficiency without mitral stenosis. Transesophageal echocardiogram was refused by the patient.

36- A 69 year old white man who is a retired shipyard worker complains of a feeling of breathlessness initially on exertion and now even at rest with significant weight loss. The patient states that he worked for over 30 years in construction, focusing primarily on the restoration of old buildings. He states the he has had a dry cough for years along with worsening fatigue and anorexia. He started smoking when he was 29 years old with an average of 1-2 packets per day.
Vitals: no fever; tachycardia (HR 115); tachypnea (RR 30); normal BP. Physical examination: clubbing; dry end inspiratory fine bibasilar crackles; loud P2, parasternal heave and JVD.
Labs- Sputum examination reveals asbestos bodies. PFTs: reduced total lung capacity, vital capacity, residual volume, and DLco; normal FEV1/FVC.

37- You are evaluating a 75-year-old woman with diabetes and hyperlipidemia complaining of shortness of breath. She reports having occasional chest pain with exertion for years, but yesterday she reported syncope with the pain. On examination, she is afebrile with mildly elevated blood pressure. Cardiac auscultation demonstrates a harsh, rasping crescendo-decrescendo systolic murmur heard best at the second intercostal space at the right upper sternal border. Her carotid pulse is small and rises slowly.

38- A 37 year old female comes to the emergency room complaining of pleuritic pain on the right side of the chest and shortness of breath together with fever and a productive cough. There is no blood with the sputum. The pain is typically sharp and stabbing and it arises when she takes a deep breath.
Upon examination the patient had decreased chest movement during inhalation on the right side; dullness on percussion of right lung base; reduced or absent breath sounds over right lung base; bronchial breath sounds

auscultated on right side; friction rib; location of dullness moves with respiration; decreased tactile fremitus over right lung.
CBC: elevated WBC count with predominance neutrophils. Gram positive diplococci on sputum smear and culture; elevated protein, decreased glucose, and many neutrophils in pleural exudate. Chest X-ray showed consolidation of right lower lobe; pleural effusion on right side.

39- A 35 year old woman presents to a cardiologist after recurrent episodes of chest pain, palpitations, and dizziness. The patient's symptoms arose while she was on a bus going to work, lasting for less than 10 minutes. She was unable to identify any specific trigger for that initial episode. She had two more similar episodes that were sufficiently intense that she thought she was going to die. Since these episodes, she fears recurrent attacks and is scared to leave the house. Other symptoms that occur during the attacks include nausea, sweating, tingling sensations, and hot flashes.
Vitals: stable. Physical and neurologic exam normal.
Labs- Lytes/CBC: normal, Vitamin B12 and folate: normal, TFTs within normal limits, urine toxicology negative, urinary catecholamines normal.
Imaging- ECG: normal sinus rhythm with no evidence of ischemia or infarction.

40- A 40 year old African American woman complains of progressive dyspnea on exertion, cough, chest discomfort, weight loss, and loss of appetite. The patient also complains of a purplish rash over her face. Her symptoms have progressed over the past year. She is a non-smoker.
Vitals: low grade fever (38.2C); tachypnea. Physical examination: mild respiratory distress; bluish-purple, swollen lesions on nose, cheeks, and earlobes; clubbing; bilateral fine inspiratory crackles.
Labs- CBC: lymphopenia. ESR elevated; hypercalcemia; hypoglobulinemia; serum ACE levels elevated. Urinalysis: 24 hour urine calcium elevated. Skin and transbronchial lung biopsies reveal noncaseating granulomas; staining and cultures negative for organisms. FEV1/FVC ratio normal.
Imaging- CXR: bilateral hilar lymphadenopathy.

41- A 63-year-old male presents to the emergency room complaining of breathlessness for the past three days. Cardiac history is positive for a myocardial infarction three years ago followed by four-vessel coronary artery bypass surgery. The patient has been asymptomatic since surgery with no complaints of chest pain. Over the last three months, the patient notes onset of shortness of breath while unloading groceries and walking stairs. Two weeks ago, he was unable to complete his daily one-mile walk at the high school track. He noted swelling in his feet and ankles. Four days prior ago, he woke at 2 am short of breath and had to sleep in his recliner the rest of the night. He has been unable to lay flat in bed at night since then and has slept on 3 pillows. Yesterday, he became breathless walking from one room to another. He presents today with extreme shortness of breath. He denies chest pain.
His past history reveals MI in 1996, CABG 4 vessel in 1996. Vague history of hypercholesteremia. Surgical history also positive for inguinal hernia repair 15 years previous. Patient smoked 1 pack/day until 1997. 30 pack/year history.
Vitals- BP 108/52; P 140, irreg.; R 30; Temp 37.2°C. Physical examination- Breathless, moderately obese male in acute distress sitting upright complaining "I am going to die. Please help me.". Neck: Distended neck veins with visible cannon waves, JVD to 12cm. Carotids without bruits. Chest: Scattered rhonchi throughout, rales bilateral one third lower bases. Cough is productive and frothy. Heart: Tachycardia and irreg. Grade 3/6 systolic murmur at LSB, S3 gallop noted. Abd: Liver palpable three centimeters below right costal margin. Hepatojugular Reflex+. Non-tender to palpation, +Bowel sounds 4 quadrants. Extremities: 4+ pitting edema of lower extremities to the knees. Nail beds minimally cyanotic, no clubbing. Pulses intact.
Lab- EKG: Left bundle branch block. Atrial fibrillation with ventricular rate of 140.
Imaging- CXR: Cardiomegaly with diffuse pulmonary infiltrate consistent with pulmonary edema.

42- A 71 year old man complains of chest pain over the last 30 minutes. He describes a severe, substernal pressure across his chest radiating to the shoulders. The pain awoke him from sleep. He also complains of nausea, fatigue, and perspiration. Vitals: normal. Physical examination: bibasilar crackles in lungs; S4 but no murmurs or rubs; weak dorsalis pedis, posterior tibialis pulses bilaterally. Labs: Cardiac biomarkers (CK, CK-MB,

troponin I): elevated. CBC: WBC mildly elevated. Imaging - CXR: mild interstitial thickening. No mediastinal widening, ECG: Acute ST elevation with involvement of proximal left anterior descending artery.

43- Sarah is a 40 year old female first presented six years ago, with breathlessness, easy fatigability and atrial fibrillation (AF). The initial echocardiogram reported moderate mitral regurgitation and normal ventricular function. She had excellent effort tolerance, was treated with digoxin and warfarin, and declined further investigation. She presented six years later with two episodes of congestive cardiac failure within a year. ECG showed AF with non-specific diffuse T- wave inversions. Chest X-ray did not show any sign of previous pulmonary tuberculosis, although the left cardiac border had an unusual double profile. Echocardiogram was repeated and this demonstrated normal left ventricular function. There was mild to moderate mitral regurgitation through a morphologically normal mitral valve and dilated mitral annulus. Bi-atrial dilatation was seen. As the history of recurrent congestive cardiac failure was incongruent with her echocardiogram findings, further investigations were arranged. Coronary angiography showed normal coronary arteries. However, a small area of linear calcification was seen near the apex during fluoroscopy. Simultaneous left and right heart studies performed after infusion of 500ml saline demonstrated elevation and equalisation of the diastolic pressures of all cardiac chambers to within 5 mmHg. The mean right atrial pressure (RAP) was 16 mmHg with a prominent Y descent, right ventricular end diastolic pressure (RVEDP) 16 mmHg, left ventricular end. The echocardiogram was repeated. This time subtle signs of diastolic checking of the posterior left ventricular wall and abnormal septal bounce were noted. Cardiac MRI showed that the left ventricle was distorted close to its apex by an indrawing of the free wall and pericardium, associated with localised calcification. This feature was responsible for the double profile seen on the left cardiac border on chest X-ray. CT thorax clearly demonstrated the band of calcification.

44- A 37 year old male CEO presents with epigastric pain and nausea. He vomited today and has a 2 week history of intermittent epigastric pain radiating to the back that worsens with eating. He works under a great deal of stress and drinks 6 cups of coffee and several drinks of alcohol each day. He also smokes 2 packs of cigarettes a day and takes aspirin on a regular basis.
Vitals: normal. Physical examination: mild pallor, abdomen soft.
Labs- CBC: hemoglobin12.1 g/dL. Amylase mildly elevated; lipase normal.

45- A 33 year old presents with progressively increasing exertional dyspnea, hemoptysis, and paroxysmal nocturnal dyspnea. She reports a childhood history of fever accompanied by join pain and skin rash; since that time she has been diagnosed with a valvular heart disease and has been receiving penicillin injections every 3 weeks.
Vitals: low volume pulse. Physical examination: mild peripheral cyanosis; left parasternal heave; loud S1 and P2; “rumbling” mid-diastolic murmur with presystolic accentuation heard at apex; opening snap heard at apex.
Imaging – CXR: left atrial enlargement; pulmonary venous congestion. Barium swallow: narrowing of the esophagus by enlarged left atrium.

46- A 76 year old man presents to the ER with chest pain, confusion, and a severe cough. The patient`s illness began with the abrupt onset of headache, muscle aches, and weakness followed 24 hours later by high fevers and shaking chills. He subsequently developed a non-productive cough with pleuritic chest pain, shortness of breath, nausea, vomiting, and diarrhea. He is a chronic smoker and drinks heavily.
Vitals: high fever (40 C); bradycardia (HR 50); tachypnea; normal BP. Physical examination: disoriented; diaphoretic; crackles bilaterally.
Labs- CBC: elevated WBC (18,000). Lytes: hyponatremia. Gram stain of sputum reveals numerous neutrophils but no bacteria, increased Legionella titers by IFA. Culture on charcoal yeast extract medium positive for Legionella.

47- A 40 year old male is brought to the ER with complaints of severe difficulty breathing and sudden onset, severe right sided chest pain. He is a chronic smoker and has predominantly emphysematous COPD.

Upon physical examination the patient had severe tachycardia, tachypnea, and hypotension with no fever. He was cyanosed with trachea shifted to the left side. Chest exam revealed hyper-resonant percussion note on right, diminished breath sounds, and decreased tactile fremitus.
ABGs revealed hypoxemia and respiratory alkalosis. ECG was normal.

48- A 69 year patient presents to his family physician complaining of chest discomfort in the epigastrium that radiates to the shoulder and jaw. The discomfort always is elicited by exertion such as climbing the stairs and relieved by rest. It lasts for approximately 3 minutes. Patient claims that he is going through a stressful situation financially and his only son abandoned him. A prior history of type II diabetes mellitus for the last 10 years which is “semi-controlled” according to patient on oral pills.
Vitals and physical examination are normal.
Labs- Cardiac biomarkers: normal. CBC/Lytes: normal.
Imaging- CXR: normal.

49- A 59-year-old male with a past medical history of COPD, alcohol abuse, and smoking is admitted to the hospital with a chief complaint of shortness of breath for 2 months. SOB is so severe that he is not able to prepare his dinner. He also reports a 18-kg weight loss during the last 6 months. The patient says that he has seen blood in his sputum. He has a 60 pack-years history of smoking.
He was admitted to the hospital 2 months ago for a weight loss work-up. A CT of the abdomen showed a 1.5 cm-pancreatic mass which did not look cancerous on ERCP. He had evidence of chronic pancreatitis. A CXR was reported as COPD changes.
Vitals; Stable. Physical examination: Chest: Markedly decreased air entry on the right, with an area of dullness. CVS: Clear S1S2.
Lab- mild hyponatremia (sodium 129 mEq/L), a hemoglobin of 15 mg/dL, and a corrected calcium level of 10.4 mg/dL.
Imaging- CXR: perihilar mass on the right. CT thorax: Heterogeneous 3.5 cm soft tissue mass encasing the right main pulmonary artery upper lobe branches and contiguous with the right main pulmonary artery superior border.

50- A 50 year old man experiences an episode of fainting while cleaning the house. Within the past 6 months, he has felt fatigue, palpitations, chest pain, and occasional shortness of breath.
Vitals: tachycardia; wide pulse pressure (BP 140/50). Physical examination: tall with long, thin limbs and high arched palate; dislocated lens in left eye; bobbing movement of head; crackles bilaterally at lung bases; displaced PMI; S3 over apex; high pitched, decrescendo, blowing diastolic murmur loudest at left sternal border; systolic blushing and diastolic blanching with gentle pressure on nail bed; water hammer pulse.
Labs- ECG: sinus rhythm with increased QRS amplitude and left ventricular hypertrophy.
Imaging- CXR: LV enlargement with cardiac apex displaced downward and to left; enlarged ascending aorta.
Echo: left ventricular hypertrophy; aortic root dilatation; flutter of anterior leaflet; early closure of mitral valve.

51- A 58-year-old woman underwent radical mastectomy for breast cancer two and half year ago. She has been complaining of low back pain and generally unwell. She developed dyspnoea and was noted to have a BP of 90/70mmHg, pulse 128/min regular. In spite of tachycardia, there was exaggerated 'X' descent on JVP.

52- A 42 year old chronic smoker and painter by occupation presents with progressively increasing breathlessness over the last 2 months. The breathlessness occurs more in the early morning hours, and there is history of allergy to dust. On physical examination the patient uses accessory muscles of respiration. On chest auscultation there are ronchi present in both lungs. Chest x-rays reveal normal findings, and ECG is also normal. Pulmonary function tests are as follows:

Test	**Patient value**	**Normal value**
FEV_1	2.2 L	3.8 L
FEV_1/ FVC	60%	77 %
Residual volume	1550ml	1200ml
Diffuse lung capacity of CO	15.9	16.0
Mid expiratory flow	4.6 L/s	4.8 L/s

53- A 55-year old male presents to your office for assessment of a chronic cough. He complains of "coughing for the last 10 years". The cough has become more bothersome lately. The cough is productive of sputum that is usually mucoid; occasionally it becomes purulent. He has a 35-year history of smoking two packs of cigarettes a day. He quit smoking approximately 2 years ago.
On physical examination his blood pressure is 160/85 mmHg. His pulse is 96 and regular. He has a BMI (body mass index) of 34. He weighs 127 kg. He wheezes while he talks. On auscultation, adventitious breath sounds are heard in all lobes. His chest x-ray reveals significant bronchial wall thickening. There are increased markings at both lung bases.

54- A 73-year-old African-American lady has had several hospital admissions over the past 2 years with the diagnosis of Congestive Heart Failure (CHF). The patient was on home oxygen therapy for CHF and Coumadin therapy for atrial fibrillation. The patient suddenly worsened with severe and progressive shortness of breath (SOB) and increasing leg edema. Her weight increased from 72.5 kg to 113 kg and she became immobile. The patient had extreme dyspnea with swelling of the legs, thighs and abdomen.
Vitals: stable. Physical examination: significantly elevated jugular venous distention of 14 cm, bilateral bibasilar crackles, reduced air entry, an irregular heart-beat with a normal S1 and a fixed split S2, a 2/6 to 3/6 holo-systolic murmur (radiating through the whole precordium and changing with respiration), a laterally displaced point of maximal impulse and a right ventricle heave. The abdomen showed significant pulsatile hepatomegaly. The extremities had 4+ edema with chronic skin changes.
Imaging- Echo: ejection fraction (EF) of 55%, paradoxical septal motion with right ventricular overload, severe right atrial and ventricular dilatation, severe tricuspid regurgitation, mild-to-moderate pulmonary insufficiency, right ventricular (RV) end-diastolic pressure of 75 mmHg and systolic pulmonary artery pressure of 80 mmHg (upper limit of normal is 40 mm Hg).
CT abdomen and pelvis: abdominal ascites, small bilateral pleural effusions with pleural scarring, focal aneurysms and dilatation of the splenic artery measuring up to 3 cm in size and dense calcification and granulomatous changes of the spleen. No evidence for chronic thrombo-embolic pulmonary disease or chronic obstructive pulmonary disease
ECG: normal sinus rhythm, incomplete right bundle branch block, T wave inversion in V1, V2, and V3 and RV hypertrophy.
Sleep study: mild sleep apnea.

55- A 30 year old at 30 weeks gestation presents to the emergency room with acute onset of left sided pleuritic chest pain and shortness of breath. First pregnancy in otherwise healthy woman. Because of a complicated prenatal course thus far, the patient has been on bed rest for 2 weeks.
Physical examination- T: 37.2 C, HR 105, shallow breathing due to pleuritic pain. Lungs are clear to auscultation. Bilateral lower extremity swelling without a palpable cord.
Labs- CBC/Lytes normal. D-dimer is elevated.
Imaging- Doppler ultrasound shows no DVT.

56- A 50-year-old chronic alcoholic, who had a roadside accident 3 days ago, presents with dyspnea, cough, and fever for the last 24 hours. Examination reveals central cyanosis, RR 28, BP 108/64 mmHg, PR 120 bpm, normal JVP, and bilateral diffuse rales. Cardiovascular examination is normal except for tachycardia. Investigations reveal bilateral diffuse alveolar infiltrates on chest x-ray, pulmonary capillary wedge pressure 10 mmHg, and PaO2 of 50 mmHg on 50% O2 inhalation therapy.

57- A 2 year old boy was noted to be pale and short of breath on a routine well child visit. The patient is a recent immigrant from Southeast Asia and has a family history of cousins dying in their first decade of life from unknown illnesses. He has had recent fatigue with occasional headaches made worse by standing.
Vitals: mild tachycardia. Physical examination: regular rhythm with a 2/6 vibratory systolic ejection murmur; conjunctiva pale with mild scleral icterus.
Labs-CBC: microcytic anemia.
Imaging- XR hand: mosaic pattern produced by trabeculation in the bones of the hand.

58- A 38–year- old woman complained of shortness of breath, low grade fever especially at night with sweats and weight loss. On examination, her blood pressure was 110/70 mmHg but JVP was raised up to the angle of mandible. Pulse was 96/min irregularly irregular with a third heat sound but no murmurs. Hepatomegaly was present and she had marked pitting edema over her leg.

59- Mike is a 72-year-old male who reports experiencing chest pain 14 days ago while he was out walking. His chest and left shoulder felt tight. He stopped walking and rested and the pain eased. He did not seek medical help at the time because he thought it was a stitch. He then experienced the same pain while out walking 3 days ago, which stopped when he sat down.

60- A 74-year old farmer was brought to the local emergency department with his wife by the county sheriff. Apparently he developed a “twinge of chest pain “while shoveling grain 3 hours ago. He insisted on staying home until “he collapsed on the floor”. Even then, he wanted to stay home and rest, but his wife insisted on calling 911. He was rushed to the emergency room (ER) in the nearest hospital. At his admission he states that the pain is “almost gone” – “what a fuss about nothing.” On taking a history he tells the admitting physician that he smokes two packs of cigarettes daily, that he drinks a “goodly amount of beer,” and that he had been told that his serum cholesterol level is good; the doctor even told him he was one of the highest values he had ever seen for a man his age. The admitting physician also notes that the patient is obese and that he seems confused on the cholesterol issue, believing that the higher the cholesterol value, the better. On further questioning he admits to a dull, aching, viselike pain around his chest, with radiation to the left shoulder. He also discloses that when the pain was at its worst he experienced nausea and vomiting, but he adds that because he already is doing better it must have been something he ate. His wife adds that she has never seen him in so much pain, but he is a “stubborn old goat.” He still insists it’s just a little stomach trouble, but on physical examination, he is sweating and diaphoretic. He has vomited twice since coming to the ER. His blood pressure is 160/100 mmHg, and his pulse is 120 and irregular. His abdomen is obese, and you believe you can detect an enlarged aorta by deep palpation.
His electrocardiogram (ECG) results reveal significant Q waves in V1 to V4 with significant ST segment elevation in the same leads. There are reciprocal ST changes (ST segment depression) in the inferior leads (II, III, avF).

61- A 38 year old man presents to the outpatient clinic with complains of shortness of breath and acute onset right sided chest pain that started yesterday afternoon. The pain and dyspnea are localized to the right side, are unrelated to position or activity, and actually started while the patient was resting. His symptoms have been stable over the past day, and he has had no progressive difficulty breathing. He has no prior medical or family history but reports that he has smoked approximately one pack of cigarettes per day for 10 years.

Vitals: mild tachycardia. Physical examination: tall and thin; in mild distress; no JVD or lymphadenopathy; diminished breath sounds, decreased tactile fremitus, and hyerresonance in right lung fields; no mediastinal or tracheal shift or cyanosis noted; cardiac exam normal.
Labs- ABGs: hypoxemia, acute respiratory alkalosis. ECG: unremarkable.

62- A 3 year old boy presents with chest pain high fever, and a productive cough. The patient has had a cold for the past week. Over the last few days the cough is worsening and associated with difficult breathing. He had several episodes of vomiting with abdominal pain on the day of admission. He was healthy prior to this episode.
Vitals: tachypnea, fever (38.3 C). Physical examination: decreased breath sounds and dullness to percussion in right lung base with rales on inspiration.
Labs: WBC= 29 (leukocytosis) with left shift.
Imaging: CXR PA/lat: right middle lobar consolidation with air bronchograms present.

63- A 33- year -old woman complains of shortness of breath, palpitation and chest discomfort. Examination revealed a thin lady, pulse regular, BP 100/60. First heart sound heard with a mid-diastolic murmur at the apex. There is a 1/4 systolic murmur at the pulmonary area and left sternal border ECG shows RBBB.

64- A 60 year old man presents to the clinic with shortness of breath, fatigue, weight loss, and night sweats. He has a 30 pack year history of smoking. He also reports difficulty swallowing and hoarseness. Vitals: tachycardic. Physical examination: engorged neck veins.
Labs- CBC: anemic. Lytes: decreased sodium.
Imaging- CT chest: large central pulmonary mass compressing esophagus.

65- A 64 year old man presents for a general medical examination. He is totally asymptomatic, has a normal physical activity but a 3/6 systolic murmur is heard over his precordium, at the base of the heart and at the apex. The murmur decreases after amyl nitrate and is unchanged after a post-extrasystolic beat.

66- A 32-year-old firefighter presents with multiple injuries after an explosion. He is gasping for air and complaining of severe lightheadedness and weakness. During the explosion, a piece of wood pierced his thoracic wall at the fifth intercostals space along the left sternal border.
Vitals: tachycardia (HR 130); hypotension (BP 90/80) unresponsive to rehydration; tachypnea. Physical Examination: cyanosis; confusion and acute distress; JVD; heart sounds distant; apical impulse diminished; inspiratory lowering of systolic BP >10 mm Hg; increased distention of neck veins during inspiration.
ECG: nonspecific ST-segment and T-wave changes; diminished QRS voltage in limb leads; QRS complexes alternating in size. CXR: mild cardiomegaly.

67- A 30-year-old man presents to the ED with an acute attack of shortness of breath, coughing, and wheezing. He states the attack occurred several hours after playing with a neighbour's cat.
Vitals: no fever; tachycardia (HR130); tachypnea (RR 40); PE: confused and diaphoretic; uses accessory muscles of respiration; cyanosis; lungs hyperresonant to percussion; inspiratory and expiratory diffuse wheezing bilaterally; increased to E-to-I ratio.
Labs- CBC: eosinophilia. ABGs: primary respiratory alkalosis with reduced PO2 and PCO2 (elevated PCO2 indicates respiratory failure); peak flows decreased. PFTs: low FEV1/FVC with >15% improvement of FEV1 following administration of B2-agonist; sputum analysis reveals Curschmann spirals (mucus that forms casts in small airways) and Charcot-Leyden crystals (eosinophil breakdown products); elevated serum IgE.
Imaging- CXR: hyperinflation, flattened diaphragms (secondary to air trapping and increased residual volume).

68- A 60-year-old man complains of progressive shortness of breath on exertion and a non-productive cough. He has a 60-pack-year smoking history.
Vitals: no fever; tachypnea (RR 24). Physical examination: moderate respiratory distress; pursed lips; using accessory muscles of respiration; barrel-shaped chest; hyper-resonant percussion note; distant breath sounds; scattered rhonchi heard bilaterally on auscultation; heart sounds distant.

Labs- ABGs; mild hypoxia with hypocapnia. PFTs: decreased FEV1/FVC ratio; decreased DLCO; increased TLC, FRC, and RV.

69- A 30-year-old patient presents with gradually worsening dyspnea and fatigue, anginal chest pain, and 2 documented episodes of pulmonary thromboemboli over the last year. The patient received a heart-lung transplant. The patient's native heart showed massive right ventricular hypertrophy. The lungs showed numerous thromboemboli, and the vasculature showed marked medial smooth muscle hypertrophy, web-like endothelial proliferations filling several arterioles, and atherosclerotic plaques on the main pulmonary arteries.

70- An 82-year-old man develops acute-onset of pleuritic chest pain and shortness of breath at rest. He has been recovering from hip replacement surgery that was performed 3 days ago. He denies a history of cardiac disease.
Vitals: tachycardia (HR 126); hypotension (BP 72/52); tachypnea (RR 32); low grade fever (38.9 C). Physical Examination: apprehensive and restless; cyanosis; localized rales, wheezes, and a pleural rub; wide physiologic splitting of S2; JVD; good peripheral pulses.
CBC: leukocytosis. Elevated ESR. D-Dimer positive. ECG: sinus tachycardia; right axis deviation and S1-Q3-T3 (flipped T wave in III) pattern seen. ABG: arterial hypoxemia, hypocapnia, and respiratory alkalosis; alveolar-arterial oxygen gradient increased.

71- A 22-year-old woman presents with mild dyspnea on exertion, arthralgia, fever and erythema nodosum. Physical examination reveals hepatosplenomegaly, generalized lymphadenopathy and corneal opacities. X-ray of the chest shows bilateral symmetrical hilar adenopathy.

72- A 60 year old man is evaluated by a physician because of weight gain and increasing abdominal girth of several months` duration, despite no change in diet. His temperature is 37.1 C, blood pressure is 70/40 mmHg, pulse is 120/min, and respirations are 18/min. Head and neck examination is remarkable for "bags" under the eyes and jugular venous distension. Chest examination demonstrates pulmonary rales and a third hear sound. Abdominal examination reveals an enlarged, tense abdomen with a palpable fluid wave; abdominal organ palpation is inadequate because of tension. Examination of the extremities demonstrates pitting edema at the ankles.

73- A 47-year-old man complains of severe shortness of breath. The patient was admitted 2 days ago for acute pancreatitis. Vitals: fever (38.5C); tachycardia (HR 112); tachypnea (RR 30-36); hypotension. Physical examination: altered mental status; central cyanosis; warm, moist skin; accessory muscles used for respiration with intercostal retractions; bilateral inspiratory rales and coarse breath sounds.
Labs- CBC: leukocytosis. ABGs: severe hypoxemia (<70 mm Hg) refractory to increased FlO2; PaO2-to-FlO2 ration less than 200:1. Swan-Ganz catheter reveals pulmonary capillary wedge pressure <18 mm Hg; elevated amylase and lipase.
Imaging- CXR: diffuse bilateral alveolar and interstitial infiltrates with normal heart size.
Echo: normal LV function.

74- A 22 year old woman is seen by a physician because she feels poorly. Physical examination demonstrates waxy pallor of her skin and mucous membranes. She also has multiple purpura on her extremities that she attributes to minor trauma, such as hitting her hand accidently on a drawer. Blood studies are performed, demonstrating a red cell count of 1.5 million/μL, white count of 1300/μL (80% lymphocytes), and platelet count of 40,000/μL. Reticulocytes are absent. All blood cells seen have normal morphology. Bone marrow biopsies obtained from the hips bilaterally show predominantly fat, which markedly diminished precursors in all blood cell lines.

75- An 11 year-old otherwise healthy girl presented to the emergency department with the sudden onset of severe, sharp mid-epigastric pain approximately 20 hours earlier. She subsequently began having episodes of nausea and vomiting, especially after trying to eat or drink. She had no hematemesis, diarrhea, melena, or hematochezia. She reported taking ibuprofen three to four times per week for five months for muscle aches or menstrual cramps. She was occasionally given an 800-milligram dose by her mother. She had last taken ibuprofen the day before the pain began. She had never been on steroids, and there was no personal or family history of endocrine problems. She had no prior surgeries.
Vitals: fever (38.7 C); pulse 120; BP 130/74. Physical examination: She was sitting up in bed with her knees flexed, avoiding sudden movements. She had diminished bowel sounds. Her abdomen was notably soft, but with exquisite tenderness in the mid-epigastrium. She was also slightly tender in the right lower quadrant. She had negative Rovsing's and obturator signs.
Labs: leukocytosis; hematocrit 40.2.

76- A 60-year-old widowed Chinese woman with a 6-month history of episodic chest tightness, shortness of breath, pain that “moves all over my body,” and numbness in her legs. These attacks, which occur once or twice weekly, occur suddenly, reaching peak intensity within a few minutes. During an attack, pain travels from her chest to her abdomen, groin, and legs. The pain is often accompanied by a sensation of intermittent “air” coming from her abdomen to her throat, making her believe that she is being choked. She also describes feeling as if she is in a closed room or small space. The patient is frustrated about her symptoms and thinks she might have a serious medical problem. She has had frequent medical evaluations by her primary care physician and second opinions from various specialists. She consulted a doctor of traditional Chinese medicine and tried some herbal medications, but has had no relief. She also has a history of a runny nose and sneezing treated by nasal corticosteroid sprays.
Vitals: normal. Physical examination: neurologic exam normal. Clear nasal discharge present bilaterally, no sinus tenderness. Labs- CBC: eosinophils elevated. Imaging- None

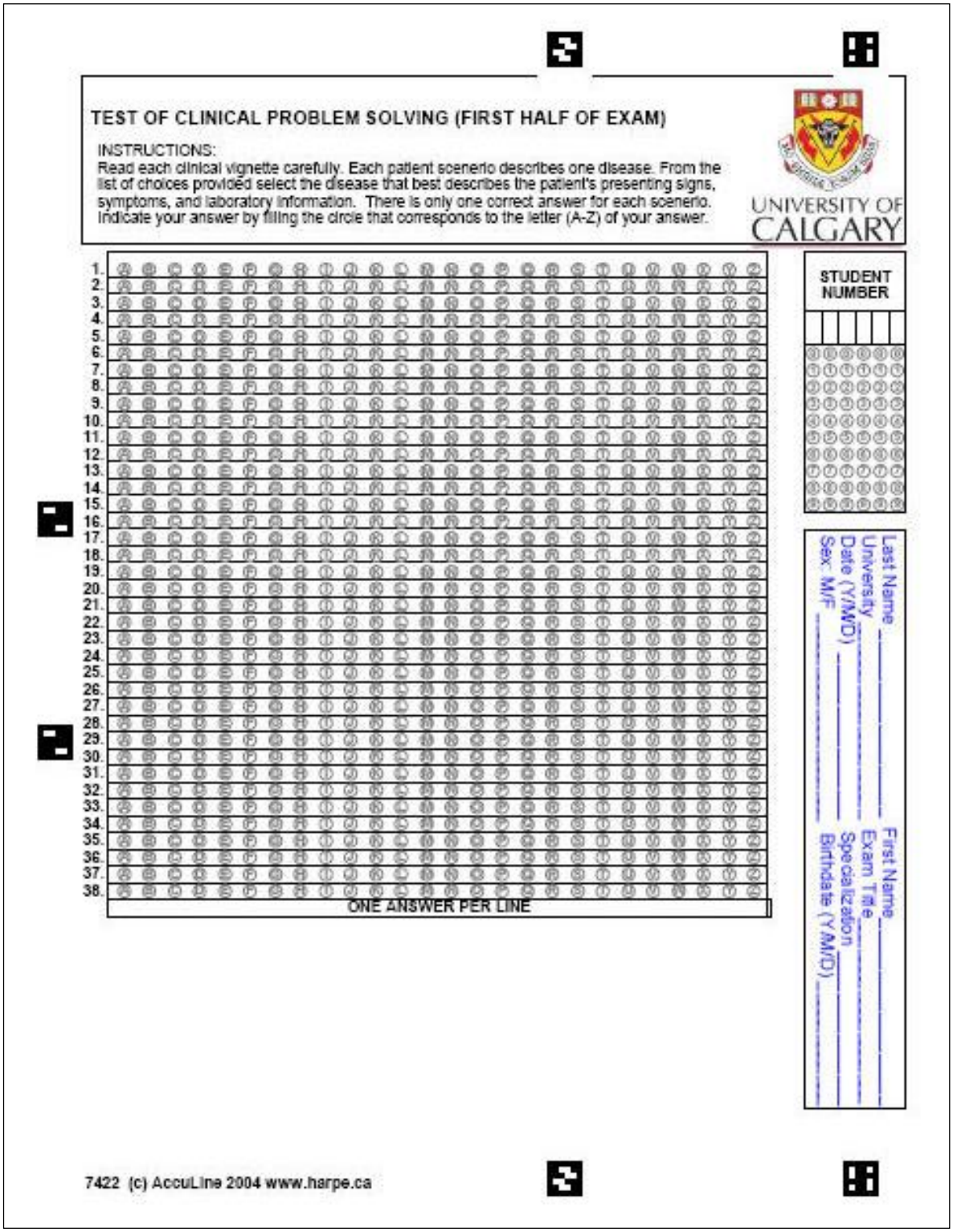
Appendix B2: Disease Options

OPTION	DISEASE	QUESTION #	CLINICAL PRESENTATION
A	Acute Coronary Syndrome	22,27,42,60	CD
B	Acute Respiratory Distress Syndrome	15,29,56,73	DYS
C	Anxiety/Panic disorder	13,14,39,76	CD
D	Anemia	10,11,57,74	DYS
E	Asthma	5,33,52,67	DYS
F	Cardiac Tamponade	9,19,51,66	DYS
G	Chronic Obstructive Pulmonary Disease	6,7,53,68	DYS
H	Congestive Heart Failure	16,34,41,72	DYS
I	Constrictive Pericarditis	1,23,43,58	CD
J	Lung Cancer	4,36,49,64	CD
K	Peptic Ulcer Disease	12,28,44,75	CD
L	Pneumonia	3,38,46,62	CD
M	Pneumothorax	2,26,47,61	CD
N	Pulmonary Embolism	8,31,55,70	DYS
O	Pulmonary Hypertension	17,18,54,69	DYS
P	Sarcoidosis	30,32,40,71	DYS
Q	Stable Angina	24,25,48,59	CD
R	Valvular Regurgitation	20,35,50,65	CD
S	Valvular Stenosis	21,37,45,63	DYS

*CD=Chest Discomfort

*DYS=Dyspnea

Appendix B3: Score Sheets



TEST OF CLINICAL PROBLEM SOLVING (FIRST HALF OF EXAM)

INSTRUCTIONS:
Read each clinical vignette carefully. Each patient scenerio describes one disease. From the list of choices provided select the disease that best describes the patient's presenting signs, symptoms, and laboratory information. There is only one correct answer for each scenerio. Indicate your answer by filling the circle that corresponds to the letter (A-Z) of your answer.

UNIVERSITY OF CALGARY

STUDENT NUMBER

Last Name
First Name
University
Exam Title
Date (Y/M/D)
Specialization
Sex M/F
Birthdate (Y/M/D)

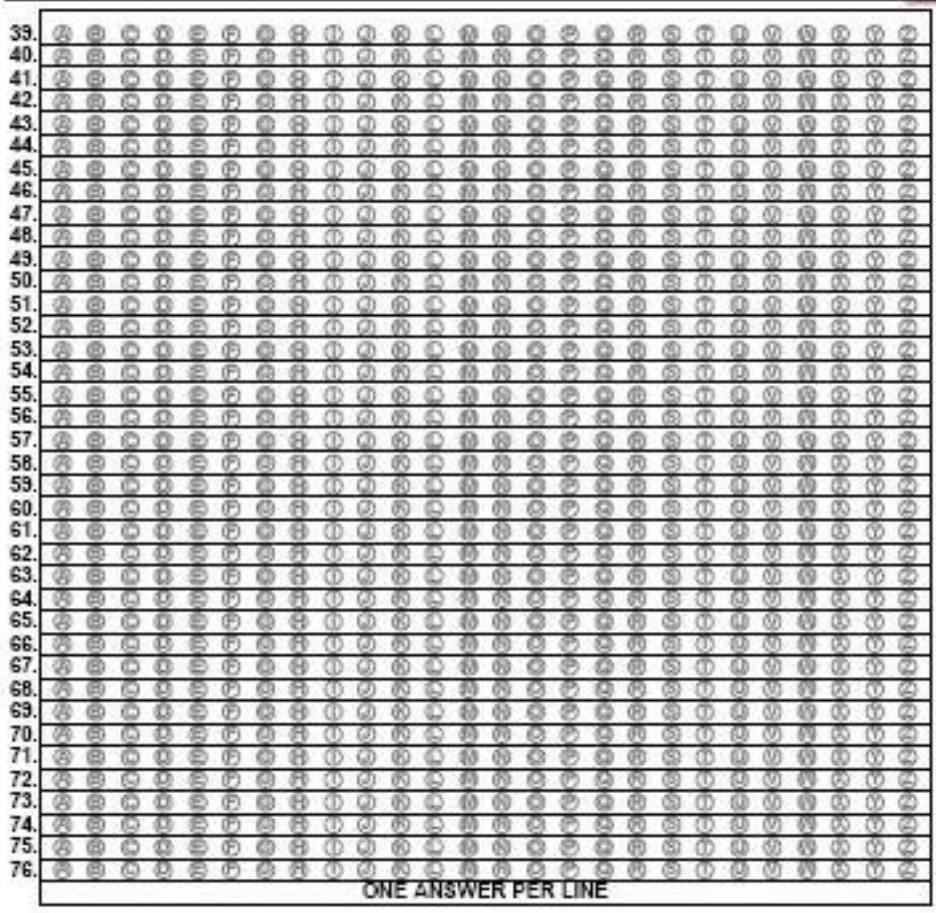
ONE ANSWER PER LINE

7422 (c) AccuLine 2004 www.harpe.ca

TEST OF CLINICAL PROBLEM SOLVING (SECOND HALF OF EXAM)

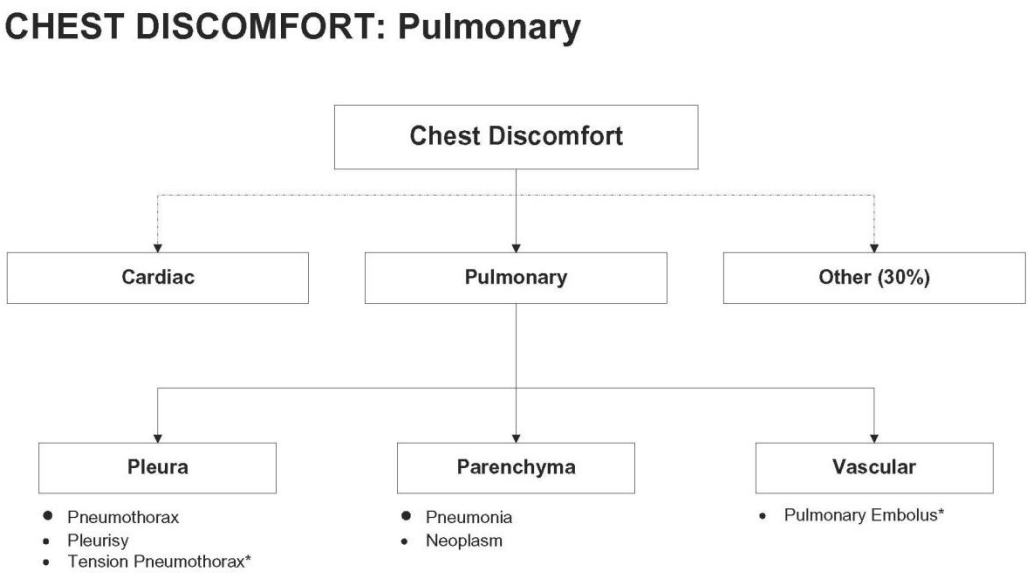
INSTRUCTIONS:
Read each clinical vignette carefully. Each patient scenerio describes one disease. From the list of choices provided select the disease that best describes the patient's presenting signs, symptoms, and laboratory information. There is only one correct answer for each scenerio. Indicate your answer by filling the circle that corresponds to the letter (A-Z) of your answer.

UNIVERSITY OF CALGARY

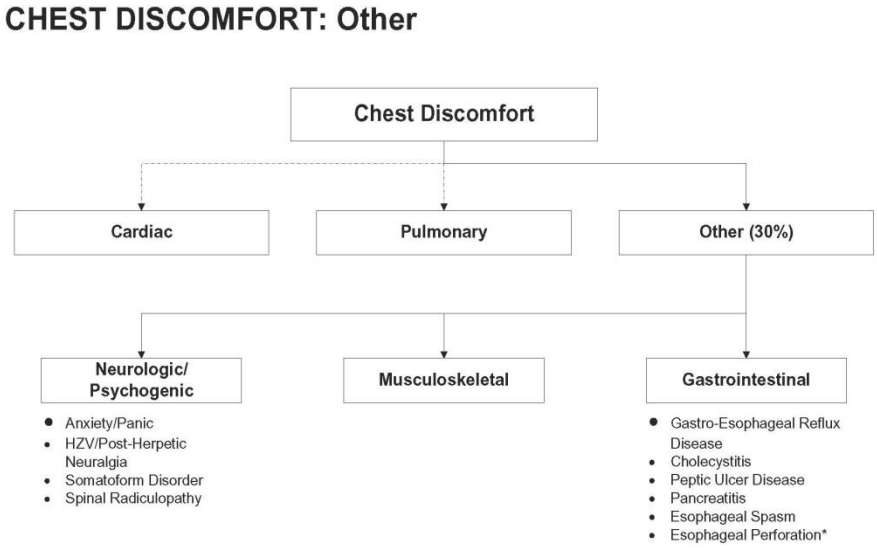


Appendix C

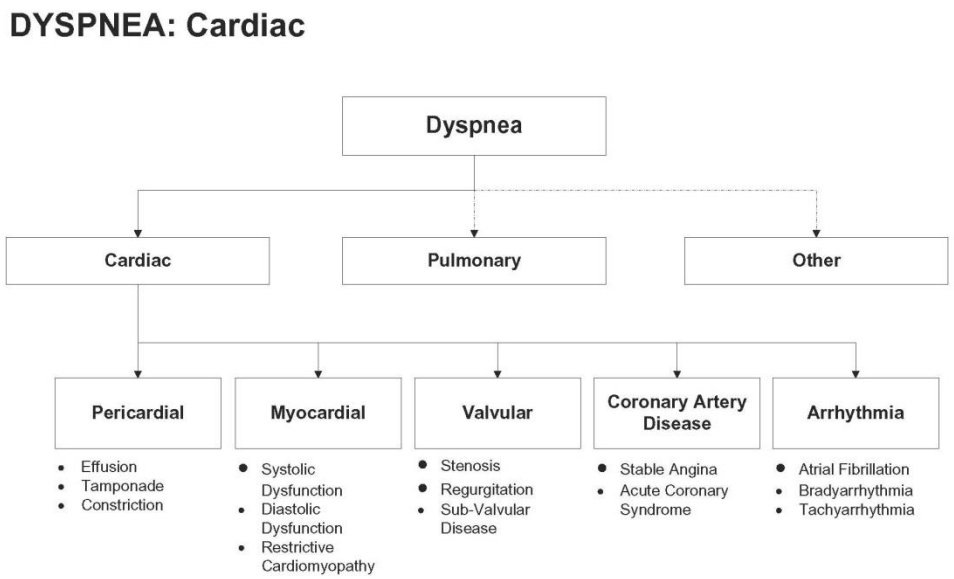
Appendix C1: Chest Discomfort Scheme-Pulmonary Causes



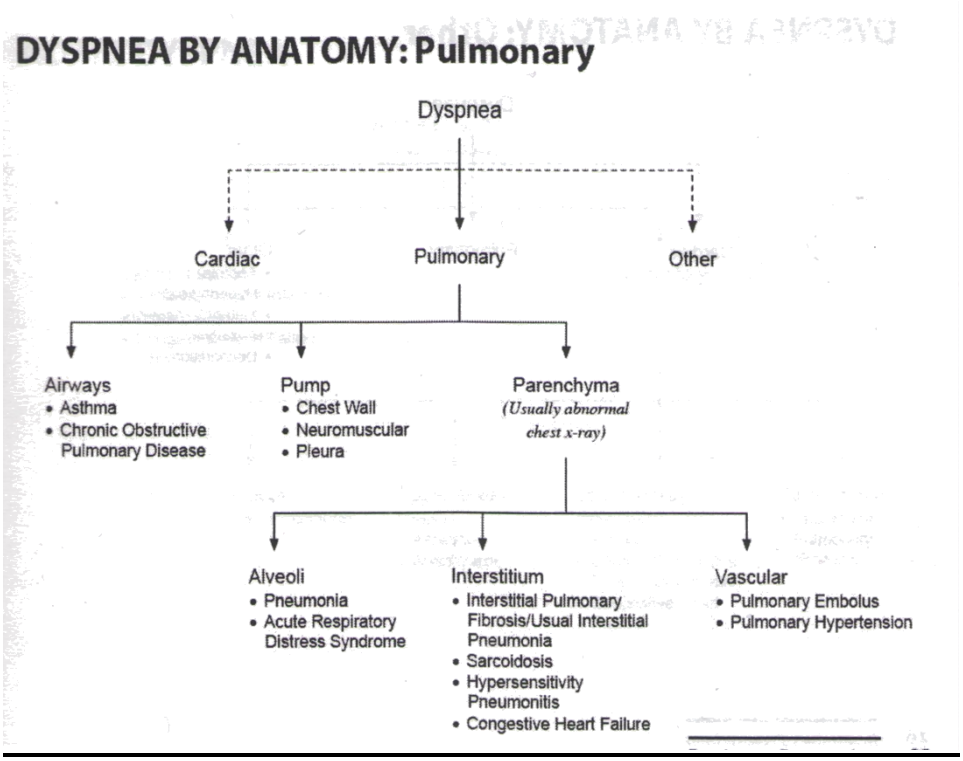
Appendix C2: Chest Discomfort Scheme-Other Causes



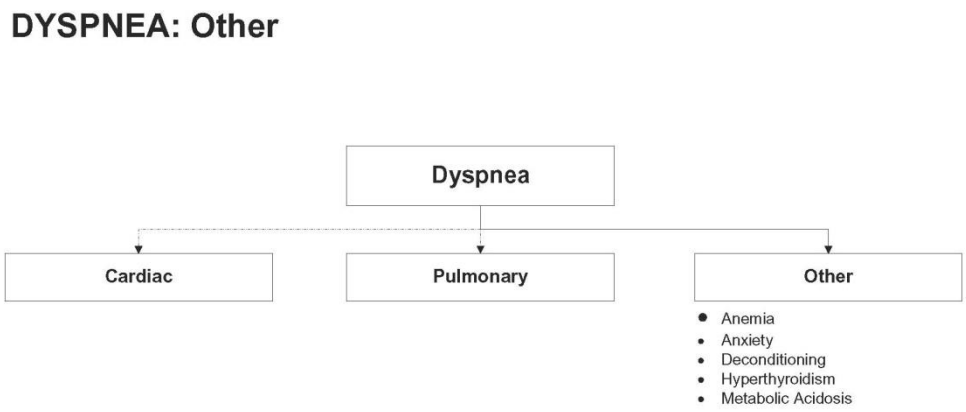
Appendix C3: Dyspnea Scheme-Cardiac Causes



Appendix C4: Dyspnea Scheme-Pulmonary Causes



Appendix C5: Dyspnea Scheme-Other Causes



Appendix D: Written Think Aloud Protocol Exercises

Task #1: Answer the following question:

A 56-year-old man presents with shortness of breath and a chronic cough productive of copious mucoid sputum. Over the past 3 years, his symptoms have occurred for at least 3 months of every year. He also reports a 40-pack-year smoking history.
Vitals: tachypnea (RR 25). Physical examination: blue lips; plethora; bilateral ronchi and wheezes.
Labs- CBC: hematocrit increased. ABGs: hypoxia and hypercapnia. ECG: presence of P pulmonale and poor progression of R wave in chest leads. PFTs: obstructive pattern of increased residual volume and decreased FEV1/FVC.
Imaging- CXR, PA: presence of increased basilar bronchovascular markings and thickened bronchial walls.

a) Congestive heart failure
b) Chronic Bronchitis
c) Asthma
d) Pneumonia

Please describe the thinking process you went through when you answered the above question and explain why you chose your answer:

Task #2: Answer the following question:

A 69 year patient presents to his family physician complaining of chest discomfort in the epigastrium that radiates to the shoulder and jaw. The discomfort always is elicited by exertion such as climbing the stairs and relieved by rest. It lasts for approximately 3 minutes. Patient claims that he is going through a stressful situation financially and his only son abandoned him. A prior history of type II diabetes mellitus for the last 10 years which is "semi-controlled" according to patient on oral pills.
Vitals and physical examination are normal.
Labs- Cardiac biomarkers: normal. CBC/Lytes: normal.
Imaging- CXR: normal.

a) Stable Angina
b) Mitral Stenosis
c) Emphysema
d) Asthma

Please describe the thinking process you went through when you answered the above question and explain why you chose your answer:

Task #3: Answer the following question:

A 10 year old boy presents with difficulty breathing, cough, and wheezing. His symptoms started 4 days ago when he contracted an upper respiratory infection. He is allergic to cats and pollen and had eczema and allergic rhinitis during infancy. His father smokes cigarettes.
Vitals: slight fever; tachypnea; tachycardia. Physical examination: unable to speak in full sentences; nasal mucosa boggy and pale; pulsus paradoxus present; moderate supraclavicular, intercostal, and subcostal retractions; decreased air movement throughout all lung fields with scattered expiratory wheezes; prolonged expiratory phase; hyperresonant to percussion.
Labs- CBC: leukocytosis with lymphocytosis and eosinophilia. Increased serum IgE. ABG: respiratory alkalosis, hypocarbia, and hypoxemia. PFTs: increased total lung capacity and residual volume; decreased FEV1/FVC.
Imaging- CXR: lung hyperinflation with flattened diaphragms; peri-bronchial cuffing and linear atelectasis in both bases.

a) Aortic Stenosis
b) Emphysema
c) Asthma
d) Pulmonary Hypertension

Please describe the thinking process you went through when you answered the above question and explain why you chose your answer:

Task #4: Answer the following question:

A 42 year old man business executive complains of recurrent dyspepsia over the past month. The patient reports epigastric discomfort that he describes as gnawing, dull, achy, intermittent, episodic, that is relieved by food or antacids. He acknowledges a high stress lifestyle with frequent business travel; he also smokes two packs of cigarettes per day and drinks heavily. Vital signs: normal. Physical examination: abdomen soft; mild tenderness to deep palpation in midepigastrium; no occult blood in stool.
Labs- CBC: hematocrit normal. Rapid urease breath test positive; serum Helicobacter pylori antibody positive; gastric mucosal biopsy reveals H. pylori infection; no evidence of malignancy.

a) Peptic Ulcer Disease
b) Aortic Stenosis
c) Aplastic Anemia
d) Pulmonary Hypertension

Please describe the thinking process you went through when you answered the above question and explain why you chosee your answer:

Appendix E: Examples of Written Think Aloud Descriptions

Clinical Presentation	Case	Written Description	Problem Solving Strategy	Diagnostic Performance
Dyspnea	Asthma	The patient has wheezing and allergy. He is an Emphysema (COPD) case.	Guessing	Wrong
Dyspnea	COPD (Chronis Bronchitis)	I don`t think its congestive heart failure because of productive copious mucoid sputum. Asthma is usually without sputum expect if with superinfection. Pneumonia is not likely because long period of symptoms. My likely diagnosis is chronic bronchitis (COPD)	Hypothetical Deductive	Correct
Chest Discomfort	Stable Angina	Cardiac markers are okay, that excludes stable angina. Asthma, no, there is no intrinsic factor for it. Emphysema, no CXR is normal These symptoms fit mitral stenosis	Hypothetical Deductive	Wrong
Chest Discomfort	Stable Angina	Patient has epi-gastric chest pain that radiates to shoulder, the pain resolves in less than 5 minutes, with poor controlled DM, this is a typical case of stable angina.	Scheme Inductive	Correct
Dyspnea	COPD (Chronis Bronchitis)	Patient has SOB and cough for 3 months for 3 years, with positive history of smoking, look cyanotic, labs shows acidosis and image finding is of an obstructive pattern. This is a case of COPD	Scheme Inductive	Correct
Chest Discomfort	Peptic Ulcer Disease	Typical peptic ulcer patient.	Pattern Recognition	Correct

Appendix F: MFRM Candidate Measurement Report

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis
68	19	3.58	3.65	191	28	1.36	.9	.97	.1	1.00	.26
66	19	3.47	3.55	171	26	.91	-.1	.75	-.4	1.18	.32
69	19	3.63	3.54	170	30	.87	-.1	.47	-1.0	1.28	.31
64	19	3.37	3.44	153	24	.76	-.5	.64	-.8	1.29	.39
65	19	3.42	3.39	146	25	.76	-.5	.58	-.9	1.25	.41
63	19	3.32	3.38	144	24	.72	-.7	.62	-.9	1.28	.38
63	19	3.32	3.38	144	24	1.35	.9	1.17	.5	.64	.16
64	19	3.37	3.33	138	24	.81	-.4	.69	-.6	1.23	.33
66	19	3.47	3.33	137	26	.82	-.3	.70	-.5	1.29	.30
62	19	3.26	3.32	137	23	1.45	1.2	1.55	1.3	.53	.17
64	19	3.37	3.32	136	24	.87	-.2	.69	-.7	1.33	.36
64	19	3.37	3.32	136	24	.41	-1.9	.50	-1.3	1.43	.37
61	19	3.21	3.26	129	23	1.10	.3	.88	-.2	1.32	.49
61	19	3.21	3.26	129	23	.43	-1.9	.46	-1.7	1.50	.44
65	19	3.42	3.25	128	25	1.02	.1	1.04	.2	.97	.23
65	19	3.42	3.25	128	25	.69	-.8	.61	-.9	1.33	.28
60	19	3.16	3.20	122	22	1.00	.1	.95	.0	1.09	.35
62	19	3.26	3.18	120	23	.87	-.2	.84	-.3	1.13	.28
59	19	3.11	3.14	115	22	.83	-.4	.84	-.3	1.23	.43
59	19	3.11	3.14	115	22	.91	-.1	.92	-.1	.92	.22
59	19	3.11	3.14	115	22	.49	-1.7	.48	-1.7	1.49	.41
59	19	3.11	3.14	115	22	.78	-.5	.72	-.7	1.33	.41
59	19	3.11	3.14	115	22	1.92	2.2	1.47	1.2	.51	.40
59	19	3.11	3.14	115	22	.74	-.7	.81	-.4	1.34	.37
61	19	3.21	3.13	114	23	.88	-.2	.95	.0	.94	.24
61	19	3.21	3.13	114	23	.83	-.4	.77	-.5	1.27	.37
61	19	3.21	3.12	112	23	.93	.0	1.09	.3	1.06	.28
63	19	3.32	3.10	111	24	.81	-.4	.78	-.4	.99	.19
58	19	3.05	3.08	109	21	.72	-.8	.91	-.1	1.26	.37
58	19	3.05	3.08	109	21	.97	.0	.86	-.3	1.15	.37
58	19	3.05	3.08	109	21	1.05	.2	1.00	.1	1.11	.38
58	19	3.05	3.08	109	21	1.01	.1	.99	.0	.79	.23
60	19	3.16	3.05	105	22	1.24	.7	1.12	.4	.76	.21
52	19	2.74	3.02	102	20	.94	.0	.96	.0	1.19	.30
57	19	3.00	3.02	102	21	.95	.0	.85	-.3	1.13	.37
57	19	3.00	3.02	102	21	1.47	1.3	1.28	.8	.54	.24
57	19	3.00	3.02	102	21	1.04	.2	1.02	.1	1.17	.41
59	19	3.11	3.00	100	22	.68	-.9	.64	-1.0	1.36	.39
51	19	2.68	2.97	97	20	.61	-1.3	.63	-1.2	1.41	.38
56	19	2.95	2.96	96	21	.56	-1.5	.57	-1.4	1.41	.37
61	19	3.21	2.95	95	23	.51	-1.6	.58	-1.2	1.26	.32
61	19	3.21	2.95	95	23	.43	-2.0	.44	-1.8	1.64	.36
58	19	3.05	2.93	93	22	.93	.0	.89	-.1	1.07	.36
58	19	3.05	2.93	93	22	.88	-.2	1.14	.5	1.18	.34
58	19	3.05	2.92	92	21	.58	-1.4	.60	-1.2	1.52	.37
58	19	3.05	2.92	92	21	1.42	1.2	1.48	1.3	.71	.27
55	19	2.89	2.90	90	20	.67	-1.0	.69	-.9	1.40	.37
55	19	2.89	2.90	90	20	.66	-1.0	.64	-1.1	1.44	.47
55	19	2.89	2.90	90	20	.72	-.8	.70	-.9	1.42	.44
60	19	3.16	2.88	88	22	.77	-.6	1.05	.2	1.10	.25
60	19	3.16	2.88	88	22	.42	-2.1	.42	-1.9	1.65	.35
60	19	3.16	2.88	88	22	1.04	.2	.88	-.2	1.22	.31
49	19	2.58	2.86	86	19	1.05	.2	1.12	.4	.89	.13
57	19	3.00	2.85	85	21	.64	-1.1	.63	-1.1	1.43	.36

Appendix F: MFRM Candidate Measurement Report (cont)

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis
57	19	3.00	2.85	85	21	.86	-.3	.91	-.1	1.19	.35
57	19	3.00	2.85	85	21	.89	-.2	.92	-.1	1.04	.23
54	19	2.84	2.84	84	20	.56	-1.5	.54	-1.6	1.57	.43
54	19	2.84	2.84	84	20	.97	.0	.98	.0	1.25	.43
54	19	2.84	2.84	84	20	1.12	.4	1.09	.3	.87	.28
54	19	2.84	2.84	84	20	.35	-2.6	.38	-2.4	1.74	.41
54	19	2.84	2.84	84	20	1.20	.6	1.09	.3	.91	.37
54	19	2.84	2.84	84	20	.99	.0	1.01	.1	1.00	.41
59	19	3.11	2.81	81	22	.89	-.2	.93	.0	1.11	.28
53	19	2.79	2.77	78	20	.62	-1.3	.65	-1.1	1.33	.39
53	19	2.79	2.77	78	20	.72	-.8	.71	-.9	1.35	.43
53	19	2.79	2.77	78	20	1.98	2.5	2.06	2.7	-.24	.12
58	19	3.05	2.74	75	21	.53	-1.6	.51	-1.6	1.49	.30
55	19	2.89	2.72	73	20	1.86	2.3	1.75	2.0	.16	.22
52	19	2.74	2.71	73	20	.77	-.7	.72	-.9	1.20	.37
52	19	2.74	2.71	73	20	.92	-.1	.89	-.2	1.22	.47
52	19	2.74	2.71	73	20	.69	-1.0	.68	-1.0	1.35	.41
52	19	2.74	2.71	73	20	1.07	.3	1.07	.3	.93	.30
52	19	2.74	2.71	73	20	.91	-.1	.91	-.1	1.26	.41
52	19	2.74	2.71	73	20	.84	-.4	.87	-.3	1.14	.32
52	19	2.74	2.71	73	20	.50	-1.9	.52	-1.7	1.72	.46
46	19	2.42	2.69	70	19	.57	-1.6	.56	-1.6	1.57	.25
46	19	2.42	2.67	69	19	.90	-.2	.87	-.3	1.13	.33
54	19	2.84	2.66	68	20	.68	-1.0	.68	-1.0	1.11	.31
51	19	2.68	2.65	67	20	.63	-1.3	.62	-1.3	1.41	.37
51	19	2.68	2.65	67	20	.90	-.2	.92	-.1	.97	.27
51	19	2.68	2.65	67	20	.57	-1.5	.66	-1.1	1.38	.33
51	19	2.68	2.65	67	20	1.83	2.2	1.81	2.2	-.15	.18
51	19	2.68	2.65	67	20	1.75	2.1	1.77	2.1	.12	.29
51	19	2.68	2.65	67	20	.97	.0	1.00	.1	1.01	.34
51	19	2.68	2.65	67	20	.61	-1.3	.59	-1.4	1.39	.34
45	19	2.37	2.63	65	19	.81	-.5	.79	-.6	1.12	.19
53	19	2.79	2.59	62	20	.97	.0	.91	-.1	1.09	.40
53	19	2.79	2.59	62	20	.80	-.6	.82	-.5	1.18	.29
50	19	2.63	2.59	62	19	.92	-.1	.94	-.1	1.00	.32
50	19	2.63	2.59	62	19	2.04	2.7	1.99	2.6	-.27	.19
50	19	2.63	2.59	62	19	1.32	1.0	1.27	.9	.56	.13
50	19	2.63	2.59	62	19	1.22	.7	1.21	.7	.75	.26
50	19	2.63	2.59	62	19	.91	-.1	.90	-.2	.93	.21
50	19	2.63	2.59	62	19	1.14	.5	1.10	.4	.87	.33
50	19	2.63	2.59	62	19	.75	-.8	.74	-.8	1.27	.34
49	19	2.58	2.53	56	19	.61	-1.4	.60	-1.4	1.55	.42
49	19	2.58	2.53	56	19	.85	-.4	.82	-.5	1.23	.41
49	19	2.58	2.53	56	19	.82	-.5	.83	-.5	1.20	.39
49	19	2.58	2.53	56	19	.56	-1.6	.57	-1.5	1.59	.45
48	19	2.53	2.47	51	19	1.04	.2	1.01	.1	.97	.39
48	19	2.53	2.47	51	19	1.03	.2	1.10	.4	.87	.29
48	19	2.53	2.47	51	19	1.38	1.2	1.45	1.4	.61	.37
48	19	2.53	2.47	51	19	1.83	2.3	1.85	2.3	-.21	.07
48	19	2.53	2.47	51	19	.67	-1.1	.66	-1.1	1.45	.44
48	19	2.53	2.47	51	19	.60	-1.4	.60	-1.4	1.60	.46
48	19	2.53	2.47	51	19	1.52	1.6	1.66	1.9	.08	.18
54	19	2.84	2.46	50	20	.77	-.6	.74	-.7	1.33	.29
42	19	2.21	2.46	50	19	1.91	2.5	1.94	2.5	-.14	.16
36	19	1.89	2.42	47	19	1.23	.8	1.19	.6	.81	.22
47	19	2.47	2.41	46	19	.35	-2.8	.36	-2.8	1.89	.44
47	19	2.47	2.41	46	19	.82	-.5	.83	-.5	1.36	.43
47	19	2.47	2.41	46	19	1.53	1.6	1.57	1.7	.40	.37
47	19	2.47	2.41	46	19	1.08	.3	1.12	.4	.80	.27
47	19	2.47	2.41	46	19	1.57	1.7	1.66	1.9	.24	.22

Appendix F: MFRM Candidate Measurement Report (cont)

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis
50	19	2.63	2.40	45	19	1.15	.5	1.13	.5	.89	.24
46	19	2.42	2.35	41	19	1.13	.5	1.12	.4	.95	.34
46	19	2.42	2.35	41	19	1.12	.4	1.12	.4	.87	.37
46	19	2.42	2.35	41	19	1.08	.3	1.06	.2	1.03	.37
40	19	2.11	2.34	40	19	1.04	.2	1.08	.3	.89	.19
49	19	2.58	2.33	40	20	3.09	4.5	3.16	4.6	-1.71	.11
45	19	2.37	2.29	35	19	.39	-2.5	.39	-2.6	1.71	.39
45	19	2.37	2.29	35	19	.67	-1.1	.67	-1.1	1.39	.35
45	19	2.37	2.29	35	19	1.14	.5	1.18	.6	.78	.35
45	19	2.37	2.29	35	19	1.45	1.4	1.40	1.3	.47	.37
45	19	2.37	2.29	35	19	.65	-1.2	.63	-1.3	1.46	.37
39	19	2.05	2.28	35	19	1.22	.7	1.25	.8	.67	.15
44	19	2.32	2.23	30	19	.60	-1.4	.63	-1.3	1.48	.37
44	19	2.32	2.23	30	19	.96	.0	.93	-.1	1.03	.21
44	19	2.32	2.23	30	19	1.05	.2	1.05	.2	1.01	.43
44	19	2.32	2.23	30	19	1.26	.8	1.24	.8	.51	.08
38	19	2.00	2.22	30	19	1.15	.5	1.09	.4	.85	.16
38	19	2.00	2.19	27	19	.45	-2.2	.46	-2.2	1.74	.44
43	19	2.26	2.17	25	19	1.28	.9	1.30	1.0	.48	.21
43	19	2.26	2.17	25	19	1.13	.5	1.14	.5	.76	.38
42	19	2.21	2.11	20	19	.89	-.2	.92	-.1	1.09	.28
42	19	2.21	2.11	20	19	.89	-.3	.86	-.4	1.15	.34
42	19	2.21	2.11	20	19	2.11	2.9	2.09	2.9	-.60	-.02
42	19	2.21	2.11	20	19	1.24	.8	1.21	.7	.83	.32
42	19	2.21	2.11	20	19	1.30	1.0	1.31	1.0	.56	.36
36	19	1.89	2.11	20	19	.71	-.9	.68	-1.0	1.32	.22
36	19	1.89	2.07	17	19	1.06	.2	1.04	.2	1.10	.40
36	19	1.89	2.07	17	19	2.70	4.0	2.67	3.9	-1.38	.20
36	19	1.89	2.07	17	19	1.62	1.8	1.63	1.8	.12	.30
41	19	2.16	2.05	15	19	1.12	.4	1.09	.4	.95	.43
41	19	2.16	2.05	15	19	.87	-.3	.90	-.2	1.09	.33
41	19	2.16	2.05	15	19	1.30	1.0	1.27	.9	.56	.27
35	19	1.84	2.01	12	19	1.32	1.0	1.31	1.0	.53	.24
35	19	1.84	2.01	12	19	2.22	3.1	2.16	3.0	-.57	.24
40	19	2.11	1.99	10	19	.92	-.1	.90	-.2	1.04	.29
40	19	2.11	1.99	10	19	.66	-1.1	.68	-1.1	1.44	.41
40	19	2.11	1.99	10	19	1.33	1.1	1.37	1.2	.45	.07
28	19	1.47	1.94	6	19	.77	-.6	.77	-.7	1.34	.22
28	19	1.47	1.94	6	19	2.37	3.2	2.49	3.5	-.72	-.01
39	19	2.05	1.94	5	19	.83	-.5	.78	-.7	1.27	.42
33	19	1.74	1.93	5	19	.76	-.7	.73	-.8	1.29	.31
33	19	1.74	1.89	1	19	1.10	.4	1.14	.5	.78	.13
37	19	1.95	1.82	-5	19	.75	-.7	.82	-.5	1.17	.17
31	19	1.63	1.81	-6	19	1.22	.7	1.20	.7	.79	.24
25	19	1.32	1.76	-11	20	.54	-1.6	.60	-1.3	1.33	.15
34	19	1.79	1.65	-20	19	.65	-1.1	.66	-1.1	1.32	.21
23	19	1.21	1.63	-22	20	.53	-1.6	.52	-1.6	1.50	.20
33	19	1.74	1.60	-25	19	.71	-.9	.66	-1.1	1.40	.38
22	19	1.16	1.56	-28	20	.87	-.3	.92	-.1	1.03	.17
32	19	1.68	1.54	-30	19	1.06	.2	1.03	.2	.90	.22
20	19	1.05	1.43	-40	21	.92	-.1	.87	-.2	1.14	.17
20	19	1.05	1.43	-40	21	1.12	.4	1.41	1.1	.65	.02
20	19	1.05	1.43	-40	21	.88	-.2	.88	-.2	1.17	.17
29	19	1.53	1.38	-46	19	.73	-.8	.72	-.8	1.23	.26
17	19	.89	1.23	-60	22	1.24	.7	1.33	.9	.73	.09
17	19	.89	1.23	-60	22	.60	-1.3	.78	-.5	1.21	.11
16	19	.84	1.16	-67	22	1.12	.4	1.05	.2	.84	.12
12	19	.63	.89	-98	25	1.17	.5	1.42	1.0	.53	.05
19	19	1.00	.84	-103	21	1.10	.4	1.01	.1	.90	.28
14	19	.74	.74	-116	24	.85	-.3	.99	.1	1.05	.34

Appendix F: MFRM Candidate Measurement Report (cont)

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis
46	19	2.42	2.35	41	19	1.12	.4	1.12	.4	.87	.37
46	19	2.42	2.35	41	19	1.08	.3	1.06	.2	1.03	.37
40	19	2.11	2.34	40	19	1.04	.2	1.08	.3	.89	.19
49	19	2.58	2.33	40	20	3.09	4.5	3.16	4.6	-1.71	.11
45	19	2.37	2.29	35	19	.39	-2.5	.39	-2.6	1.71	.39
45	19	2.37	2.29	35	19	.67	-1.1	.67	-1.1	1.39	.35
45	19	2.37	2.29	35	19	1.14	.5	1.18	.6	.78	.35
45	19	2.37	2.29	35	19	1.45	1.4	1.40	1.3	.47	.37
45	19	2.37	2.29	35	19	.65	-1.2	.63	-1.3	1.46	.37
39	19	2.05	2.28	35	19	1.22	.7	1.25	.8	.67	.15
44	19	2.32	2.23	30	19	.60	-1.4	.63	-1.3	1.48	.37
44	19	2.32	2.23	30	19	.96	.0	.93	-.1	1.03	.21
44	19	2.32	2.23	30	19	1.05	.2	1.05	.2	1.01	.43
44	19	2.32	2.23	30	19	1.26	.8	1.24	.8	.51	.08
38	19	2.00	2.22	30	19	1.15	.5	1.09	.4	.85	.16
38	19	2.00	2.19	27	19	.45	-2.2	.46	-2.2	1.74	.44
43	19	2.26	2.17	25	19	1.28	.9	1.30	1.0	.48	.21
43	19	2.26	2.17	25	19	1.13	.5	1.14	.5	.76	.38
42	19	2.21	2.11	20	19	.89	-.2	.92	-.1	1.09	.28
42	19	2.21	2.11	20	19	.89	-.3	.86	-.4	1.15	.34
42	19	2.21	2.11	20	19	2.11	2.9	2.09	2.9	-.60	-.02
42	19	2.21	2.11	20	19	1.24	.8	1.21	.7	.83	.32
42	19	2.21	2.11	20	19	1.30	1.0	1.31	1.0	.56	.36
36	19	1.89	2.11	20	19	.71	-.9	.68	-1.0	1.32	.22
36	19	1.89	2.07	17	19	1.06	.2	1.04	.2	1.10	.40
36	19	1.89	2.07	17	19	2.70	4.0	2.67	3.9	-1.38	.20
36	19	1.89	2.07	17	19	1.62	1.8	1.63	1.8	.12	.30
41	19	2.16	2.05	15	19	1.12	.4	1.09	.4	.95	.43
41	19	2.16	2.05	15	19	.87	-.3	.90	-.2	1.09	.33
41	19	2.16	2.05	15	19	1.30	1.0	1.27	.9	.56	.27
35	19	1.84	2.01	12	19	1.32	1.0	1.31	1.0	.53	.24
35	19	1.84	2.01	12	19	2.22	3.1	2.16	3.0	-.57	.24
40	19	2.11	1.99	10	19	.92	-.1	.90	-.2	1.04	.29
40	19	2.11	1.99	10	19	.66	-1.1	.68	-1.1	1.44	.41
40	19	2.11	1.99	10	19	1.33	1.1	1.37	1.2	.45	.07
28	19	1.47	1.94	6	19	.77	-.6	.77	-.7	1.34	.22
28	19	1.47	1.94	6	19	2.37	3.2	2.49	3.5	-.72	-.01
39	19	2.05	1.94	5	19	.83	-.5	.78	-.7	1.27	.42
33	19	1.74	1.93	5	19	.76	-.7	.73	-.8	1.29	.31
33	19	1.74	1.89	1	19	1.10	.4	1.14	.5	.78	.13
37	19	1.95	1.82	-5	19	.75	-.7	.82	-.5	1.17	.17
31	19	1.63	1.81	-6	19	1.22	.7	1.20	.7	.79	.24
25	19	1.32	1.76	-11	20	.54	-1.6	.60	-1.3	1.33	.15
34	19	1.79	1.65	-20	19	.65	-1.1	.66	-1.1	1.32	.21
23	19	1.21	1.63	-22	20	.53	-1.6	.52	-1.6	1.50	.20
33	19	1.74	1.60	-25	19	.71	-.9	.66	-1.1	1.40	.38
22	19	1.16	1.56	-28	20	.87	-.3	.92	-.1	1.03	.17
32	19	1.68	1.54	-30	19	1.06	.2	1.03	.2	.90	.22
20	19	1.05	1.43	-40	21	.92	-.1	.87	-.2	1.14	.17
20	19	1.05	1.43	-40	21	1.12	.4	1.41	1.1	.65	.02
20	19	1.05	1.43	-40	21	.88	-.2	.88	-.2	1.17	.17
29	19	1.53	1.38	-46	19	.73	-.8	.72	-.8	1.23	.26
17	19	.89	1.23	-60	22	1.24	.7	1.33	.9	.73	.09
17	19	.89	1.23	-60	22	.60	-1.3	.78	-.5	1.21	.11
16	19	.84	1.16	-67	22	1.12	.4	1.05	.2	.84	.12
12	19	.63	.89	-98	25	1.17	.5	1.42	1.0	.53	.05
19	19	1.00	.84	-103	21	1.10	.4	1.01	.1	.90	.28
14	19	.74	.74	-116	24	.85	-.3	.99	.1	1.05	.34
15	19	.79	.64	-131	23	1.37	1.0	1.67	1.6	.52	.17

Appendix G: MFRM Disease Measurement Report

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Nu	Disease
202	175	1.15	1.02	147	7	1.19	1.7	1.16	1.4	.79	.28	10	Acute Respiratory Distress Syndrome
254	175	1.45	1.33	115	6	1.06	.6	1.05	.5	.91	.30	12	Pulmonary Hypertension
304	175	1.74	1.62	88	6	1.03	.2	1.10	.9	.88	.25	1	Constrictive Pericarditis
311	175	1.78	1.66	84	6	1.23	2.1	1.32	2.8	.58	.20	6	Cardiac Tamponade
315	175	1.80	1.68	82	6	1.03	.3	1.00	.0	.99	.39	13	Valvular Regurgitation
396	175	2.26	2.19	38	6	.54	-5.5	.56	-5.3	1.44	.39	11	Congestive Heart Failure
427	175	2.44	2.41	19	6	1.06	.6	1.15	1.4	.76	.27	3	Asthma
448	175	2.56	2.53	8	6	1.08	.8	1.03	.3	1.00	.44	14	Valvular Stenosis
467	175	2.67	2.66	-3	6	1.13	1.2	1.08	.7	.84	.34	16	Stable Angina
481	175	2.75	2.76	-12	7	1.09	.9	1.06	.6	.94	.37	4	COPD
505	175	2.89	2.92	-27	7	1.15	1.4	1.08	.7	.98	.39	17	Sarcoidosis
508	175	2.90	2.93	-28	7	1.02	.2	.96	-.3	1.09	.40	19	Pneumothorax
508	175	2.90	2.94	-29	7	.85	-1.4	.80	-1.9	1.28	.45	5	Pulmonary Embolism
521	175	2.98	3.02	-37	7	.93	-.6	.92	-.7	1.10	.44	15	Acute Coronary Syndrome
527	175	3.01	3.07	-42	7	.94	-.5	.91	-.8	1.08	.39	7	Anemia
530	175	3.03	3.09	-44	7	.93	-.5	.91	-.8	1.09	.36	2	Lung Cancer
557	175	3.18	3.26	-64	7	1.03	.2	1.05	.4	1.00	.33	9	Anxiety/Panic disorder
567	175	3.24	3.32	-72	8	.92	-.6	.88	-.9	1.15	.41	18	Pneumonia
589	175	3.37	3.46	-91	8	.90	-.8	.82	-1.2	1.10	.37	8	Peptic Ulcer Disease
443.0	175.0	2.53	2.52	7	7	1.01	.0	.99	-.1		.36	Mean (Count: 19)	
111.0	.0	.63	.71	66	0	.15	1.6	.16	1.7		.07	S.D. (Population)	
114.0	.0	.65	.73	68	1	.15	1.6	.17	1.7		.07	S.D. (Sample)	

Appendix H: MFRM Clinical Problem Solving Approach Measurement Report

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Exact Agree. Obs %	Exact Agree. Exp %	N	Clinical Problem Solving Approach
637	447	1.43	1.98	56	4	1.02	.3	1.05	.7	.94	.22	.0	.0	1	Guessing
6288	2414	2.60	2.63	-1	2	1.03	.9	1.01	.3	.98	.32	.0	.0	2	Hypothetical Deductive
1492	464	3.22	2.99	-34	5	.88	-1.6	.83	-2.2	1.14	.26	.0	.0	3	Scheme Inductive
2805.7	1108.3	2.42	2.54	7	4	.98	-.1	.97	-.4		.27				Mean (Count: 3)
2487.0	923.3	.74	.42	37	1	.07	1.1	.09	1.3		.04				S.D. (Population)
3045.9	1130.8	.91	.52	46	2	.08	1.4	.12	1.6		.05				S.D. (Sample)