INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

THE ASSESSMENT PROCESS OF PSYCHOLOGISTS AS A FUNCTION OF CLINICAL EXPERIENCE

Leanna J. Zozula

Department of Psychology

and

Cognitive Studies in Medicine: Centre for Medical Education,

Faculty of Medicine

McGill University,

Montreal

A thesis submitted to the Faculty of Graduate Studies and Research

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

An immense thank you is due to the thirty-five subjects who participated in this research, many of whom were busy professionals. I am deeply indebted to them for sharing their valuable time with me.

The final acknowledgement is to the agencies that have helped to fund this project. I would like to thank the agencies of FCAR (Fonds pour la Formation de Chercheurs et L'Aide à la Recherche) and NSERC (National Science and Engineering Research Council) for their financial support throughout my graduate career. A last thank you goes to McGill University for providing funds through an internal grant that was instrumental in directly funding this project.

ABSTRACT

The purpose of this research was to examine the assessment process in clinical psychology using an ecologically valid design. By capitalizing on the methodologies of cognitive science and adapting data analytic techniques borrowed from the medical decision-making literature, the clinical assessment behaviours of psychology undergraduate students, clinical psychology graduate students, and professional psychologists were compared.

The results demonstrate that professional psychologists took longest to develop their diagnostic formulations, divide their questioning evenly between close and open-ended questions, provide more information to the patient, provide diagnoses that are fully consistent with the clinical data, and demonstrate an increase in diagnostic accuracy using the MultiAxial Form (APA, 1994). Finally, extensive clinical experience affords clinicians with a greater degree of case comprehension that may impact the quality of treatment.

Graduate students in clinical psychology demonstrate the same degree of test familiarity as the psychologists, adhere to empirical recommendations by conducting comprehensive clinical interviews and asking about DSM-IV criteria during the clinical interview, evidence a reliance on open-ended questions during the clinical interview, and also show an increase in diagnostic accuracy using the MultiAxial Form (APA, 1994). There were no significant differences in accuracy rates between professional psychologists and graduate students; however, some graduate students included diagnoses that were inconsistent with the clinical data.

The undergraduate students tended to cover many of the same topics in the interview as the other groups; however, they did so by posing significantly more close-ended questions. Further, none of the undergraduate students provided an accurate diagnosis of the case. There were no differences in confidence ratings across levels of experience. A model of clinical assessment behaviour is proposed that can function as a framework for future studies.

PRESENTATIONS

Parts of the work presented in this thesis have been reported at the following scientific meetings:

Zozula, L. & Patel, V.L. (1999). Age as an Influence on Evidence Use in Diagnostic Decision-making. Presented at the 4th Annual Centres of Excellence HEALNet "Putting Health Research to Work" conference. March 28 to 31, 1999. Calgary, Alberta.

Zozula, L. & Patel, V.L. (1999). Clinical Decisions by Psychologists as a Function of Clinical Practice. Twenty first Annual Conference of the Cognitive Science Society, August 19 to August 22, 1999 in Vancouver, BC.

TABLE OF CONTENTS

CHAPTER TITLES

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

What brings you here to see me today?

Subject 23A
Professional Psychologist
11 years clinical experience

This general inquiry, in its various forms, is expressed by psychiatrists, psychologists, clinical social workers, nurses, physicians and other healthcare professionals. It marks the beginning of a helping relationship, and invites the patient, in a non-directive manner, to articulate what is of concern to them (Morrison, 1995). However, the goals of the person asking this question may differ. For example, a clinical psychologist who is initializing a comprehensive assessment of the patient may be motivated by differing goals and objectives than a clinical psychologist who is beginning psychotherapy with a patient, and is armed with the results of such an assessment.

Psychological testing, psychological assessment and psychotherapy are professional activities that clinical psychologists engage in. As Meyer and colleagues (2001) specify, psychological testing is a process wherein a particular scale, instrument, or test is administered to obtain a specific score, index or profile. In contrast, they define psychological assessment as being

> concerned with the clinician who takes a variety of test scores, generally obtained from multiple test methods, and considers the data in the context of history, referral information, and observed behavior to understand the person being evaluated, to answer the referral questions, and then to communicate findings to the patient, his or her significant other, and referral sources. (Meyer et. al., 2001, p.143)

Psychotherapy, in contrast to psychological testing or psychological assessment, is a method of treatment of mental disorders and behavioural disturbances using psychological techniques and principles.

According to a practitioner survey of clinical psychologists, assessment of patients is considered to be second only to psychotherapy in terms of professional importance (Phelps, Eisman & Kohout, 1998). The ability to conduct psychological assessments, as defined above, is unique to clinical psychologists. An example of the most widely used, and basic assessment method is the clinical interview, typically unstructured in nature, beginning with a similar statement to the one found at the start of this thesis. At the end of the assessment process, a comprehensive diagnostic picture emerges, often including one or more diagnoses.

A review of published studies on psychological assessment reveals that there are plenty of studies investigating the inaccuracy of psychodiagnosis and the fallibility of clinical judgment. Comparatively, there is a lack of studies investigating the actual assessment process in psychology. There is an unmet need for not only the examination of the process that leads up to the diagnostic formulation in clinical psychology, but designs that incorporate realistic tasks so that approximations to actual clinical behaviour can be captured. Such an endeavor would provide much needed empirically derived information to inform those who conduct psychological assessments, resulting in improvements in assessment skills and techniques. Thus, this thesis seeks to examine the assessment process in clinical psychologists, with particular emphasis placed on the clinical interview, use of testing material and the development of diagnostic formulations.

1.1 State of Knowledge in Psychiatry and Psychology

A comprehensive assessment and subsequent diagnosis is often the first step in successfully treating patients with medical conditions as well as those with psychiatric and psychological disorders. Diagnosis in medicine as well as in psychiatry and psychology is a method of classification. The ideal diagnostic nomenclature should serve to communicate information subsumed by the diagnostic

label, optimize prognostic considerations, explain the etiology and guide choice of treatment (Zubin, 1978). The main difference between a medical diagnosis and a psychiatric or psychological diagnosis is that psychiatric problems rarely consist of signs, but rather are a list of symptoms, often referred to as diagnostic criteria (Goodwin & Guze, 1996; APA, 1994). Signs of a disease or disorder are objective entities such as plaque formation in the brain, a rash or a fever, all of which are physiological manifestations. Symptoms are a patient's complaints about their thoughts, feelings and behaviour. The challenge in basing psychiatric diagnoses on symptoms is that behaviours change over time and so do patient self-reports about their thoughts and feelings (Goodwin & Guze, 1996). Further, our understanding as to the mechanisms of psychiatric or psychological disorders is comparatively meager, and as Goodwin & Guze (1996) simply state "....for most psychiatric conditions there *are* no explanations" p.xiii.

The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV, APA, 1994), is one of two internationally accepted standards for nomenclature and diagnosis in psychiatric practice. The ICD-10 is used in many parts of the world, while the DSM-IV is the most widely used and accepted diagnostic tool for the classification of mental disorders in North America (Dunne & Chute, 1999). The DSM-IV defines a mental disorder as:

> A clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above. (APA, 1994, pp. xxi-xxii)

Characteristic of the fields of psychiatry and clinical psychology is the absence of clearly delineated boundaries between mental health and mental disorders. Once mental health is found to be lacking, one encounters difficulties navigating the vague distinctions between various mental disorders. Taxonomic difficulty arises due to the inherent difficulties in classifying and diagnosing mental disorders as well as unascertained etiologies for most mental disorders (Meehl, 1999; Zubin, 1978). The DSM-IV (APA, 1994) relies on phenomenological data to group mental disorders into discrete categories. While it is not based on any particular theoretical framework, it is generally agreed that a medical model is implicit (Barron, 1998; Ivey & Ivey, 1999). Critics have pointed out that basing diagnoses solely on phenomenology is of limited clinical utility because etiology, course, treatment response, biological precursors, familial and genetic influence are all potentially useful as well (Follette & Houts, 1996). This state of knowledge of disorders and psychopathology in clinical psychology contrasts with our understanding of the state of knowledge in the field of medicine. As Patel, Evans and Kaufman (1989) write, a coherent epistemological framework is essential to the examination of the medical decision-making process, with a developed understanding of the structure and content of medical knowledge. This level of understanding regarding the structure and content of clinical knowledge is lacking in clinical psychology.

Despite these limitations in the current classification system and our understanding of psychopathologies, psychologists continue to assess and diagnose patients in clinical practice. Motivated by the challenge of investigating diagnoses in clinical psychology, this thesis seeks to understand how diagnostic decisions are made in everyday practice. In particular, this thesis examines the assessment and diagnostic process in clinical psychology as a function of clinical experience from no training to experienced professional. Further, it explores the effect of extensive clinical experience, as acquired through decades of practice, on the assessment process. The effect of experience as a factor in the assessment process is examined because of the documented role practice and the accumulation of vast amounts of knowledge, both of which can increase with increasing experience in a particular domain, can have on the performance of individuals in their specialties (Ericsson,

1996). In addition, characteristics of participants who provided an accurate diagnosis will be outlined, and a model of clinical behaviour proposed.

1.2 Research objectives

The research objectives of this thesis are:

1) To examine the assessment process in psychology as a function of clinical experience. Experience in this context plays a development role from a beginner in the field of clinical psychology to an experienced and seasoned professional.

2) To examine characteristics of participants who provided an accurate diagnosis.

3) To develop a model that accurately captures the clinical assessment process in psychology. A model such as this may serve to advance a framework for future studies of the assessment process.

4) To explore the effect of extensive clinical experience on the assessment process. Experience in this context plays a maintenance and refinement role in terms of professional abilities.

1.3 Thesis organization

The thesis is conceptually organized into the following chapters. The preceding Chapter presented background information on the problem domain of assessment in clinical psychology, as well as outlining the primary objectives of this work. Chapter II surveys the literature on the nature of expert performance in general, the classification of mental disorders and limitations to this system, a selected review of clinical decision-making in clinical psychology, and illustrative findings from medicine are presented. Chapter III submits some theoretical and methodological considerations at the start, and continues to outline the methods, procedures and modes of analysis employed in the study, along with examples and excerpts of the coding schemes. In Chapter IV, qualitative as well as quantitative results of the

study are presented, with emphasis on group differences associated with levels of experience, and group differences associated with accurate diagnosis. A model of the assessment process is also proposed in Chapter IV. The final chapter, Chapter V, concludes with a brief summary and discussion of top level findings of the thesis, limitations to the study, and concluding remarks.

& Fischoff, 1988). The psychological studies of decision-making began by basing investigations on normative (or "rational") models formulated outside of psychology (Montgomery & Svenson, 1989). The early studies concentrated on how people's judgments agree with normative requirements, focusing on the decision choice. Then, during the late 1950s, Herbert Simon's work had a profound influence on the psychological study of decision-making. He sharply criticized the notion of *expected utility maximization*, which describes decision-making behaviour as maximizing the best possible outcome, given the uncertainty of a particular situation. Simon (1956) argued that actual decision-making behaviours are not normative (or "rational"), but that human cognition is limited in comparison. His term *bounded rationality* meant that human decision-making is not rational in the normative sense, because thinking this way requires excessive cognitive effort. Instead, humans make decisions by *satisficing,* that is attempting to attain a satisfactory level of achievement, by using a simplifying decision strategy or heuristic. His conceptualization of human decision-making highlighted the need to look at perception, cognition, and learning in decision-making studies, and prompted researchers to take more of an information-processing view (Slovic, Lichtenstein & Fischoff, 1988). In this view, the examination of psychological processes, knowledge representations and use of information is the primary target of investigation, rather than accuracy of the decision. Although this study does examine diagnostic accuracy, it was also designed to examine the use of information in the development of diagnostic formulations.

2.2 The Nature of Expert Performance

What does it take to become an expert in a field? Are we born with a particular ability or is it acquired, or a combination of both? What enables an expert to perform well? The study of expertise seeks to understand "what distinguishes outstanding individuals in a domain from less outstanding individuals in that domain, as well as from people in general" (Ericsson & Smith, 1991, p.2).

Chase and Simon (1973) proposed a general theory for the structure of expertise and this *original expertise approach* attempts to describe the critical performance under standardized conditions, to conduct a careful analysis of the attained performance, and to identify components of the performance that make it superior (Ericsson & Smith, 1991). Using the expertise approach, researchers have examined expert performance in many domains and have found some important characteristics of experts' performance, which are robust and generalizable across the various domains that have been studied.

2.2.1 Characteristics of Experts

Experts have demonstrated various characteristics that, taken collectively, can describe their behaviours and qualities across areas of specialization (Glaser & Chi, 1988; Shanteau, 1988). Some of the relevant qualities with respect to this thesis are described below.

The first broad characteristic is that experts excel mainly within the constraints of their own domains. The reason for this is that experts have a great deal of domain specific knowledge (Schneider, 1996). The nature of this knowledge is limited to the boundaries of their domain of expertise, therefore, knowing much and performing expertly in one domain leaves little time, energy and carry over effect for use in another domain.

A second characteristic of experts is that experts perceive large meaningful patterns in their domain (Trotter, 1986). This ability to see meaningful clusters of information does not imply a generally superior perceptual ability, but it more accurately reflects a hierarchical organization of an extensive knowledge base, resulting in superior pattern recognition (Shanteau, 1988). They are able to extract information that non-experts either overlook or are unable to see.

Experts have a sense of what is relevant and irrelevant when making decisions (Herl, O'Neil, Chung et. al., 1999). Although consistently detecting the relevant from irrelevant is difficult, experts are better at this than novices. Related to this

concept is the fact that experts have an ability to simplify complex problems, breaking them down into manageable pieces (Glaser & Chi, 1988).

Another quality of experts is that they see and represent a problem in their domain at a deeper (more principled) level than novices, while novices tend to represent a problem at a superficial level (Kuipers & Kassirer, 1984). The conceptual categories contained in the problem representations of experts are semantically organized, while the conceptual categories of novices are syntactical, or surface-feature based (Hardiman, Dufresne & Mestre, 1989).

Experts also spend a great deal of time analyzing a problem qualitatively (Schraagen, Chipman & Shalin, 2000). Experts typically try to comprehend or understand the nature of the problem. While spending the extra time understanding, they build a mental representation from which they can infer relationships among concepts and problem constraints (Glaser & Chi, 1988).

In difficult situations in their domain of expertise, experts are able to handle adversity better than non-experts (Shanteau, 1988). Even when the situation is difficult, experts continue to make effective decisions. A related concept is that experts are able to work well under stressful conditions (Dino, Shanteau, Binkley, & Spenser, 1984).

Another characteristic of experts is that they know how to learn from past decisions and to make appropriate changes in future decision strategies (Shanteau, 1988). Experts tend to be more responsive to past successes and failures, whereas novices frequently rationalize or defend past decisions, rather than learning from them. Experience per se is not sufficient to produce expertise (Brehmer, 1980); what is important is what is learned from that experience (Kolodner, 1984).

Although expert decision makers may make small errors, they generally avoid large mistakes. Especially when time is limited, the first available viable solution is selected for action, rather than a comparison of alternatives (Klein & Claderwood, 1991). For most decisions, experts have generally discovered that coming close is often good enough. Apparently, the key is not to worry about being exactly right, but to avoid making poor decisions. Experts are likely to use a two pronged strategy

by first coming up with a "ball park" estimate, and then conducting a more careful analysis (Shanteau, 1988).

Finally, experts usually use a "divide and conquer" strategy (Shanteau, 1988). Experts break large, complex, difficult problems into constituent parts, work on solving those parts, then put those partial solutions back together again, thus systematically separating decisions into parts that are more manageable.

Thus, the study of expertise places emphasis on understanding characteristics of experts, their cognitive reasoning strategies and abilities, and their knowledge structures. Research into the professional domains of expertise, such as medicine, psychology, physics and business are domains that are knowledge-rich, requiring long periods of preparatory education and training (Ericsson, 1996). Due to the nature of these professions, problems can be ill-structured, and the reliability of superior performance, even among experts, can be compromised. Expert professionals from these knowledge-rich domains do not exhibit consistently superior performance on relevant activities in their domain, especially when tests are performed under standardized conditions (Camerer & Johnson, 1991; Ericsson, 1996), thus creating the need to investigate experts in these professionals using a more naturalistic approach. To date, much of the research on clinical decision-making in psychology focused on the failure of experts to consistently perform at high levels, and the failure of experts to predict better than statistical models. However, it is important to keep in mind that the lack of consistent superior performance of experts is not particular to the field of clinical psychology, but is characteristic of other professional, knowledge-rich domains, with ill-defined problems. Examples include computer programming (Doane, Pellegrino, & Klatzky, 1990), mathematics (Lewis, 1981), physics (Reif & Allen, 1992), investing (McClosky, 1990) and medicine (Kassirer & Gorrey, 1978).

2.2.2 Choice of Terms: Experience over Expertise

Robyn Dawes (1994) defines expertise "in terms of what experts accomplish, not in terms of how they go about their task" (Dawes, 1994, p.82). As far as defining

accomplishments and measuring expert performance in any domain "few reach the highest levels of achievement and performance" (Ericsson, 1996, p.1). Given that experts in any field are understood to exhibit outstanding and superior human performance, there arises a difficulty in selecting true experts from a o ng competent individuals in a field. In chess, expertise is quantified using ELO ratings that rank individuals relative to their tournament scores. In clinical psychology, it could be possible to define an expert as someone with an advanced degree (Ph.D. or Psy.D.) with at least ten years post graduate clinical experience. However, one would be sampling professionals with experience, rather than expertise per se. As Dawes (1994) and Ericsson (1996) suggest, expertise should be based on superior, not just competent, performance in the expert's domain. Using these criteria, relatively few people in any domain should warrant being called an expert. To ere on the side of caution, this thesis uses the term experienced professional rather than expert per se. Nevertheless, many of the experienced professionals sampled in this study can be expected to exhibit numerous expert characteristics as outlined above.

Before an examination of previous research on assessment in clinical psychology can be undertaken, an understanding of the classification system currently in use is essential to appreciating the state and nature of knowledge we have of psychopathology and mental disorders.

2.3 The Classification of Mental Disorders

It has been stated that the four major goals or steps in our understanding of science are description, measurement, prediction, and control (Bingham, 1923; Jaffe, 1998). These goals or steps can be conceptualized as being hierarchical in nature. According to these four hierarchical goals of science, the first basic step in any scientific endeavor is a description of the phenomenon of interest. Once we have amassed a number of critical observations about the scientific phenomenon, we summarize them into constructs, models and theories, then apply these generalizations to specific cases in order to predict (Hempel, 1965). For mental disorders, the first step, description, is analogous to classification.

According to Maxmen and Ward (1995), there are two major approaches to the classification of diagnostic psychopathology. The first is descriptive in nature, whereby diagnoses are based on relatively objective phenomena such as signs, symptoms and natural history that require minimal clinical judgement. The second approach to diagnostic psychopathology is psychological in nature, whereby diagnoses are based primarily on inferred causes and mechanisms. The psychological approach also contemplates descriptive phenomena, but the focus is on the underlying forces behind the phenomena rather than superficial manifestations. They sum up by stating that "the descriptive approach focuses on the *what* of behavior, the psychological on the *why*" (p.8).

The DSM-I (APA, 1952) was the first official nomenclature for psychiatric disorders. It emphasized psychological (primarily psychodynamic) etiologies in the terminology and the diagnoses were loosely defined. In trying to be flexible, DSM-II (APA, 1968) contained diagnostic categories that were vague, idiosyncratic and susceptible to bias. Essentially, both the first and second editions of the DSM suffered from low interrater reliability, and poor validity (Spitzer, Forman, & Nee, 1979). The categories did not define disorders as having predictable symptoms, natural histories or responses to treatment. To address these limitations, researchers devised explicit, readily verifiable, and specific diagnostic criteria, culminating in the publication of DSM-III (APA, 1980). The DSM-III also introduced the multiaxial system, so that one could underscore the distinction between the more florid Axis I mental disorders and the chronic but subtle Axis II personality (Maxmen & Ward, 1995). As more scientific evidence accumulated, a revised edition the DSM-III-R (APA, 1986) and finally the current edition of the DSM, the DSM-IV (APA,1994) were published.

2.3.1 Limitations of the Current Diagnostic System: DSM-IV

The following brief history and introduction to the current diagnostic classification system for mental disorders helps to guide evaluation of research findings in clinical assessment.

How well a diagnosis defines a disorder and guides treatment depends on its reliability and validity. Reliability requires that a diagnostician arrive at the same diagnosis every time with little error or that different diagnosticians agree on the diagnosis. Reliability is one prerequisite to validity of a diagnostic category. Traditionally, one establishes the validity of measurement by reference or comparison to a gold standard. In psychiatry and psychology, there is usually no observable gold standard to reference (Holzer, Nguyen & Hirschfeld, 1996). Critics of DSM-IV have outlined numerous limitations with the latest DSM version (Clark, Watson, & Reynolds, 1995; Folette & Houts, 1996), including issues of poor validity.

The first major criticism of the DSM-IV is that diagnostic classification is based on a categorical rather than dimensional approach. The fact that DSM-IV is based on a categorical approach to diagnosis (APA, 1994), implies an assumption that mental illness is discontinuous with normal behaviour. As in medicine, this type of approach works best for disorders that have homogeneous characteristics and clear boundaries differentiating them from other disorders (Maxmen & Ward, 1995). The categorical approach does not work well with disorders that exist on a continuum - those that can be better described by dimensions have a greater variability within the diagnostic classification system. Especially with psychopathology and mental illness, almost every disorder has dimensional as well as categorical aspects. Even DSM-IV (1994) itself states that "there is no assumption that each category of mental disorder is a completely discrete entity with absolute boundaries dividing it from other disorders" (p. xxii). Clark, Watson and Reynolds (1995) outline undesirable consequences of the categorical approach to diagnosing psychopathology using the DSM-IV, including an elevated incidence of comorbidity, increased within-diagnostic-category heterogeneity, and the frequency

of Not Otherwise Specified (NOS) diagnoses. APA (1994) felt it impossible for the diagnostic nomenclature to cover every possible clinical presentation. Therefore, each diagnostic class, such as psychotic disorders, or mood disorders, contains at least one diagnosis of Not Otherwise Specified (NOS), which can ensure 100% coverage of psychopathology.

The second major criticism of the current diagnostic system has to do with its supposed atheoretical nature and the proliferation of diagnostic categories. Hempel (1965) proposed a model of scientific progress whereby scientific advances occur when it eventually becomes necessary to synthesize an organizing theory that describes the fundamental principles underlying the taxonomy, resulting in a reduction in the number of taxonomic categories. According to Hempel's model (1965), there appears to be little evidence that DSM is making this type of scientific progress, for the difficulty in having an atheoretical system slows research and scientific progress (Folette & Houts, 1996). The modern DSMs have claimed to be atheoretical, but any successful classification system entails some level of theory. Failing to specify the theory causes basic definitional problems that limit the utility of the classification system.

A final criticism of the DSM-IV is that "although empirical considerations were weighted especially strongly in the latest revision, many decisions were made on the basis of expert consensus in the absence of data" (Clark, Watson & Reynolds, 1995, p.147). Essentially, some diagnoses were included in the DSM-IV based on committee vote.

Advantages of the DSM-IV include reasonable diagnostic reliability, especially when structured clinical interviews are used in the diagnostic process (Holzer, Nguyen, & Hirschfeld, 1996). In addition, it recognizes that most patients with the same mental disorder do not have identical clinical characteristics. Diagnoses require some core criteria, but offer a choice among others. Finally, it was extensively field tested before publication, indicating that most of the diagnostic categories and labels are based on empirical evidence.

2.4 Clinical Decision-making in Psychology

In research on decision-making, there are two approaches in the investigation of expert performance. According to Camerer and Johnson (1991), the first view emerges from behavioral research on decision-making. This view tends to be skeptical about experts. Data from this approach to studying expert performance demonstrate that experts are not categorically better predictors than less expert counterparts (Dawes, 1994). Furthermore, this view recommends a simple mathematical model incorporating linear regression procedures for replacing the decisions of experts due to the superior accuracy of these linear combination models (Kleinmutz, 1990). Critics argue that this behavioral approach does not mimic expert decision-making strategies, but the focal point of this approach is that the regression models make consistently more accurate decisions than experts do (Camerer & Johnson, 1991).

The second approach to expert decision-making stems from research in cognitive science. Research from the cognitive science perspective suggests that expertise is a rare skill that develops only after extensive training, practice and experience (Ericsson, Krampe & Tesch-Romer, 1993). The cognitive strategies and processes of experts are more sophisticated than those of novices in the field and this elevated degree of sophistication is presumed to produce better predictions. Models generated from the cognitive science perspective attempt to mimic the decision-making strategies of experts, resulting in "expert (or knowledge-based) systems". These models attempt to emulate, rather than exceed the performance of the expert is represents.

In sum, these two approaches have different goals: the behavioral approach emphasizes the *performance* of experts, whereas the cognitive science approach emphasizes differences in experts' *processes* (Johnson, 1988). In the study of clinical decision-making in psychology, there is a comparative overabundance of studies using the behavioral, or emphasis on accuracy of performance, approach, as opposed to the cognitive science approach. Much of the studies have compared expert judgment to statistical models of varying degrees of sophistication. Other

approaches include comparisons of experts with novices, or comparisons of experienced experts with inexperienced experts in the realm of medicine. This thesis seeks to make comparisons along the continuum of experience (latter two approaches) rather than comparison with mathematical models.

2.4.1 Clinical Judgment and Predictions

It should be stated at the outset that much of the pioneering groundwork in the study of judgment and prediction in clinical psychology was sparked by Meehl's (1954) influential book, *Clinical Versus Statistical Prediction*. The ensuing debate in the literature carried on for years (Einhorn, 1986) but has subsided in intensity as of late. This explains the preponderance of earlier citations used in the following literature review.

In the 1950s, Meehl (1954) reviewed the empirical literature containing comparisons of clinical and statistical prediction methods. He compared accuracy rates of clinicians to simple linear prediction models, and concluded that the models either predicted with the same degree of accuracy or better than the clinicians. Thus, Meehl set the foundation for research into clinical judgment with a focus on investigating diagnostic accuracy and putting into question the validity and accuracy of clinical prediction.

In response, Holt (1958) attempted to refute Meehl's conclusions by pointing out that the comparisons were unfair in that they pitted what he termed "naïve clinical prediction" with "pure actuarial prediction". Naïve clinical prediction is clinical intuition without the benefit of test results, in other words no other sources of input. Pure actuarial prediction is a statistical method to combine data void of clinical judgment input.

Some time later, Sawyer (1966) criticised Meehl and his subsequent reviews of the literature on the grounds that clinical prediction (data combination) cannot be evaluated independently from its source (measurement). So Sawyer employed an eight-fold classification of prediction methods as a basis for reanalyzing the results of the published literature on statistical versus clinical prediction (Meehl

are combined (interpretation) for the purposes of clinical prediction. Clinicians usually, but not always, have access to a great deal of relevant information, but it is the intuitive manner in which this information is combined - rather than statistical - that is the empirically weaker method. Given the same information, the clinician is unable to improve on mathematically optimal methods of data combination, and may even be unlikely to approximate the accuracy of statistical prediction. Proponents for the clinical method believe that psychologists are privy to information that a computer does not have, such as formal/informal contacts with the patient, friends, family, or from other psychological forms of data. Critics of the clinical method argue that they are skeptical regarding the ability of clinicians to process large amounts of input data, they are skeptical of the predictive value of such additional data and that clinicians can code this same additional information to be processed more accurately by the computer.

As for research investigating the effect of expertise or experience on diagnostic accuracy, Goldberg (1959) determined that inexperienced practitioners performed as well as experienced ones, indicating that clinical experience does not affect or increase judgmental accuracy. The amount of information one works with can affect diagnostic accuracy. In fact, increasing amounts of test data do not necessarily lead to increasing amounts of reliability and validity in psychodiagnosis (Golden, 1964). However, too little clinical information produces poor reliability. The relationship between the amount of information available to the clinician and the accuracy of judgments is complex rather than linear, and it varies according to the particular data source made available to him or her (Wiggins, 1973). It is possible that increasing amounts of information may improve diagnostic accuracy to a certain point, but beyond that, increasing amounts of information may impair and interfere with diagnostic accuracy. It should be noted that it is the nature of the clinical information (quality and usefulness) that can improve diagnostic decisions, rather than simply the amount.

Finally, Oskamp (1962) in his study determined that the most confident clinicians tended to be the least accurate ones. Further, additional information relating to the case tends to inflate clinical confidence. Finally, he also concurred

with Goldberg (1959) in that experience alone does not increase predictive accuracy.

Despite the consistency of many of the previously mentioned findings regarding the superiority of statistical methods of data combination over clinical, or intuitive methods, there have been criticisms of the research. Dawes (1994) lists some of the major objections to empirical results demonstrating the relative inferiority of psychologists' judgment and prediction abilities. Included among these are 1) that the studies were methodologically flawed 2) that the people studied were not truly experts 3) there were no well defined expert tasks which were specifically developed to sample expert behaviour and 4) the predictive tasks presented to the clinical experts were not ecologically valid. Although some of these points may be valid ones (although Dawes (1994) systematically refuted these major criticisms), it is difficult to ignore research that has consistently demonstrated that clinical judgmental accuracy is "not systematically related to level of training, experience, the amount of information available to the clinician, or the clinician's discipline" (Faust & Nurcombe, 1989, p.197).

Interestingly, despite the findings reviewed in this section, humans continue to use their own judgment and prediction abilities. Although several explanations can be proposed for this (Kleinmuntz, 1990), the fact remains that data combination and data interpretation in psychology and psychiatry are conducted primarily by humans – not by computers - and this process should be evaluated in its own right, as opposed to simply comparing its accuracy to statistical models.

2.4.2 Cognitive Heuristics and Biases

Motivated to provide explanations as to the poor predictive abilities of humans in general and experts of any field in particular, researchers in decision-making focussed their attention on looking at how humans make predictions and decisions, and how these decisions can be improved. The most famous contribution came from Tversky and Kahneman (1974) who proposed three types of heuristics that characterize human judgment in situations of uncertainty. They are the

Representative Heuristic, the Availability Heuristic and the Adjustment and Anchoring Heuristics, and examples of each follow below.

In clinical psychology, the Representativeness Heuristic may occur when clinicians try to determine probabilistically whether a patient belongs to a certain diagnostic group. The bias occurs when this probability is assessed by the degree to which a patient is representative of, or similar to, the stereotypic patient of that group (Tversky and Kahneman, 1974). It has been demonstrated that this heuristic best describes how diagnoses are made in clinical psychology (Garb, 1998). Typically, a clinician compares the patient to a "prototypical" or representative patient that belongs to a particular diagnostic class. If the patient is similar to the prototypical one, then the diagnosis is made.

The Availability Heuristic occurs when clinicians try to judge the frequency of a class or the probability of an event based on the ease with which certain instances or occurrences can be brought to mind (Tversky and Kahneman, 1974). For example, in making a diagnosis, one may be more likely to make a diagnosis of Hypochondriasis than Somatization Disorder if one can more easily recall patients they've seen who have had Hypochondriasis. This can also lead to errors because the availability of patient information (ease with which one can recall this information) is affected by factors such as salience, recent occurrences and familiarity.

Adjustment and Anchoring Heuristics occur when different starting points yield different responses or conclusions (Tversky and Kahneman, 1974). For example, if a judgment or diagnosis changes depending on whether a piece of information or test result is known early or later on in an assessment or interview, then anchoring and adjustment is said to occur.

Reliance on heuristics can lead to biases in problem-solving and eventually erroneous decisions (Garb, 1998). The emerging judgment and decision literature is attending increasingly to debiasing, which is aimed at identifying variables that contribute to poor judgment, in hopes of controlling and eliminating systematic bias. These include identification of mistakes and fallacies, structured and systematic

ways to collect data, and maintaining critical thinking skills (see Gambrill, 1990, for a book dedicated to overcoming limitations in clinical decision-making).

2.5 Medical Expertise from a Cognitive Science Perspective

In recent years, much attention has focussed on medical expertise from a cogniti science perspective. Thus, the diagnostic reasoning strategies of novices, intermediates and medical experts have been well-documented (Schmidt & Boshuizen, 1993; Patel, Arocha & Kaufman, 1994). In medicine, novices are beginner medical students, whereas intermediates are advanced medical students and/or residents, who have knowledge of a particular field of medicine but they do not have extensive clinical experience (Patel & Groen, 1991). Experts are typically board-certified physicians with at least ten years experience. The common findings are that novices process information at a superficial level, and therefore search strategies into possible etiologies and diagnoses are quite limited. Conversely, intermediates engage in extraneous search, and they provide extensive elaborations in explaining a patient's symptomology (Arocha, Patel & Patel, 1993). Medical experts' knowledge is "finely tuned" so that they are able to hone in on critical items and to filter out irrelevant information, preventing an extensive search (Patel, Arocha & Kaufman, 1994). They very efficiently recognize, focus and elaborate on the essential components of a patient's health profile, due to their comprehensive knowledge as well as their history of clinical exposure and experience. Typically, medical experts conduct what is referred to as a situational assessment (Klein & Calderwood, 1991), where the formulation of an accurate diagnosis depends on understanding the details, parameters and circumstances under which the patient currently manifests his/her symptoms. This is an essential elaboration before a diagnosis is provided, relatively early on (Kushniruk, Patel & Fleiszer, 1995). Finally, a last characteristic of medical expert performance is the high level of accuracy and intricate comprehension of the patient's condition and diagnosis.

This thesis seeks to utilize a cognitive science approach to the investigation of diagnostic reasoning and clinical decision-making in psychology and to adapt

CHAPTER III

RESEARCH METHODOLOGY

3.1 Theoretical and Methodological Considerations

3.1.1 Naturalistic Framework

Traditional models of decision-making have failed to adequately capture the decision-making process in complex, ill-structured environments (Klein, 1993). The study of naturalistic decision-making is an attempt to rectify this situation by outlining a framework that examines the process of complex decision-making. Further, naturalistic decision-making seeks to investigate the process in real world environments, rather than laboratory settings that may artificially simplify the decision process.

According to Klein and Calderwood (1991), prevailing paradigms in decision research have limited utility in domains characterized by high degrees of uncertainty, ambiguity, ill-defined goals, time pressures, continually changing conditions and decision responsibilities that are distributed among several individuals. Research has suggested a metaphor of a decision tree where the decision maker is faced with "moments of choice" or "alternatives" that can be represented as branches emanating from a single point in this decision tree, with the decision event as the focus of research concern (Raiffa, 1968). This decision tree paradigm of research has found that human decision-makers are inherently biased and suboptimal (Klein & Calderwood, 1991). Consequently, many training programs, decisions aids and procedures have been developed to debias judgments. These analytical approaches turn out to not be very useful or effective in real world environments because the designs from which the data are based are not representative of what actually occurs

in the real world setting – especially in complex domains (Klein and Calderwood, 1991). This stresses the importance of incorporating a naturalistic approach to the investigation of decision-making, especially when developing methods to ameliorate performance in a given field. If an adequate understanding of decision processes is not achieved, aids and decision support systems that are developed without such knowledge will not work.

This present investigation seeks to study the complex problem of assessment and diagnosis in clinical psychology by using, among others, approaches borrowed from the naturalistic decision-making literature. These were specifically developed for domains that are knowledge rich and characterized by a high degree of ambiguity, uncertainty and ill-defined goals (Klein & Calderwood, 1991). Further, the sequence of tasks given to subjects was designed to mimic patterns seen in clinical practice. Specifically, psychologists in clinical practice usually receive a referral note, or telephone call outlining basic demographics and a sentence or two about presenting difficulties. Next, psychologists typically interview the patient, and may follow the interview up with psychological testing. The information derived from the assessment process (interview and testing) is typically used to formulate a provisional diagnosis and in the case of the treating psychotherapist, a tentative treatment plan (Morrison, 1995). This ordering of events was conserved in order to maintain a relative naturalistic approach to investigating the assessment process in clinical psychologists.

3.1.2 The Assembly of Existing Methods

In order to examine decision-making and problem-solving during the assessment process in clinical psychology, research methods from medicine, with its comparatively more extensive knowledge base and understanding of disease states, were adapted to enable a focus of both process and outcome oriented variables during the assessment process in clinical psychology.

The ability to accurately capture diagnostic decision-making may be dependent upon the method of analysis (Patel & Groen, 1986). Think Aloud protocols (Ericsson & Simon, 1993), where subjects verbalize their thoughts as they problem solve, combined with Propositional Analysis techniques (Van Dijk & Kintsch, 1983; Frederiksen, 1975) and the construction of knowledge networks extracted from the protocols (Patel & Groen, 1986), have proven to be useful methods of analysis in understanding the nature of medical decision-making.

These methods of analysis were adapted for use in this study due to differing task structures and knowledge bases between medicine and clinical psychology. The following serves as a brief overview of what methods were used, while the details of how these methods were used in this study can be found in Section 3.3, Methods of Data Analysis.

The Think Aloud method (Ericsson & Simon, 1983), as is the case in medicine, was not employed during the doctor-patient interaction, or clinical interview, although it was used in the explanation-based sections, such as diagnostic formulation. The rationale behind employing concurrent rather than retrospective think aloud methods is that concurrent methods provide an accurate account of the thinking processes of the subjects, while retrospective methods provide an account of what the subject thought they were thinking at the time (Kuusela & Paul, 2000). Retrospective accounts are limited due to the limits and fallibility of human memory. Concurrent think aloud protocols during the clinical interview functions to interfere with the nature of the psychologist-patient interaction, and decreases the ecological validity of the process (Patel, Evans & Kaufman, 1989). Instead, specialized coding schemes specific to psychology were developed to capture the type of knowledge acquired during this interaction.

Propositional Analysis techniques (Van Dijk & Kintsch, 1983; Frederiksen, 1975; Patel & Groen, 1986) are used in conjunction with transcribed verbal material. Typically, the verbal material, or protocol, is broken down into propositional units, which are meant to represent the lowest level of interest. The levels of propositional analysis vary depending on the goals of the study, which guide the determination of meaningful units of analysis. For this study, the meaningful unit for the interview

interaction was a psychologist-patient question response pairing. Coding criteria specific to clinical psychology were developed to analyze the interview section of the assessment process. For the think aloud diagnostic formulation section, the meaningful units of analysis were broken down into short phrases, based on syntactic cues such as sentence clauses or natural pauses in speech. These segments were used to produce diagnostic networks for each subject.

Networks are representations of what underlies the "running" of a diagnostic process (deKleer & Brown, 1983). The construction of diagnostic networks from the protocols (Patel & Groen, 1986) permit analyses of the diagnostic representations of subjects in this study. Please refer to Section 3.3 for details and an example of how this was accomplished.

3.1.3 Rationale for Simulation

The purpose of this study is to examine the assessment process in clinical psychology. Several possibilities for the presentation of clinical material were considered, including presenting subjects with a videotape of a patient or written case material. This would have provided subjects with very consistent data, in that each subject would receive the same clinical information in the same order and format. However, this option limits our ability to capture knowledge about the dynamic nature of the assessment process, namely what information is sought as relevant, how clinicians ask questions, in what order, and so on. The best choice for the purposes of this study was to opt for a simulated patient approach, which involves high fidelity simulations where the conditions of a clinical interview are reproduced in detail (Patel, Evans & Kaufman, 1989).

The researcher decided to play the role of the patient in a rehearsed and reliable manner, which bypassed issues of privacy, sensitivity of material and reluctance to come forth as a mental health patient if a real patient simulation was used. Patient simulation by the actual researcher has been done before in at least one study investigating medical diagnoses (Kassirer & Gorry, 1978), but not in studies investigating decision-making in psychology. To maximize fidelity of responses using this approach, responses to questions were well rehearsed and were based on a

script. Particular attention was paid to assure consistency in responding with regard to non-verbal actions such as stuttering (presenting complaint), tone of voice, posture and willingness to cooperate, as these can be used as clinical information. A pilot study with four subjects ensured an adequate degree of practice before the study participants were tested. Finally, portrayal of the researcher as a credible simulated patient was confirmed with validated clinical instruments that determined consistent diagnoses across instruments (see section 3.1.6 below on Determination and Conceptualization of Accuracy). For similar approach that used an actual patient in the simulation, see Patel, Evans, and Kaufman (1989).

3.1.4 Validity Issues

There has been a recent focus on examining the ecological validity of experiments designed to study decision-making and problem solving (Vicente & Wang, 1998). Reliably superior performance by experts is exhibited under conditions that capture those activities that are central to expert performance in a particular domain (Ericsson, 1996). Given the complexities of professional domains such as medicine and psychology, there exists a challenge to investigate expert performance. The difficulty is that individual experts rarely encounter the exact same problem or case. Further, standardized tasks such as fixed descriptions fail to capture the essence of the domain by eliminating the analysis of ongoing and dynamic interaction between professional and patient and between other professionals (Ericsson, 1996). Further, past research into the study of clinical decision-making of psychologists has been criticized for the use of artificial methods and tasks, which do not represent real life behaviour of clinical psychologists.

This study seeks to improve upon past investigations by emphasizing an ecologically valid, or realistic approach by asking subjects to adhere to their usual interviewing techniques, and structuring the diagnostic task as it exists in clinical practice. This permits the capture of dynamic interaction of the psychologist and patient. Again, this was part of the rationale behind selecting a simulated patient approach over videotaped or written clinical case material. Included in the measures

given to subjects was a rating scale asking them to rate the ecological validity of the study, so that perspectives from reviewers of this study as well as participants in the study can be compared.

3.1.5 Selection of a Diagnostic Problem

The selection of a diagnostic problem for this study was a critical step in ensuring that the objectives of the study be met. It was necessary that the problem case meet two essential criteria before it could be considered for portrayal in this study. The first consideration was that the assessment and diagnosis of the patient must have been challenging and complex to provide variability in performance that can be systematically examined (Patel, Groen, & Arocha, 1990). A prototypical patient suffering from a straightforward disorder would not have enabled much of a comparison of strategies between subjects of differing levels of experience. The patient material would have been confirmatory to the diagnosis, with no complicating or contradictory evidence. This rarely occurs in clinical practice, therefore, a representative patient was sought to capture the challenging nature of assessment in clinical psychology.

The second consideration was that the details of the case must be quasi-realistically portrayed by the researcher who was a female in her twenties, to ensure ecological validity and believability of the diagnostic task.

A senior and well-respected clinical psychologist in the Montreal area (E. G.) was consulted to aid in the development of a clinical case. The final selection for the problem case was based on an actual patient whom the consulting psychologist assessed and treated for a number of years, thus providing the researcher with ample amounts of rich information about the patient. The primary diagnosis this patient received was Borderline Personality Disorder, with a previous diagnosis of Anorexia Nervosa when she was an adolescent.

According to the DSM-IV (APA, 1994), a personality disorder is an "enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has an onset in

adolescence or early adulthood, is stable over time, and leads to distress or impairment" (APA, 1994, p.629). The Axis II personality disorders are grouped into three clusters based on behavioral similarities. The Cluster A Personality Disorders are Paranoid, Schizoid, and Schizotypal Personality Disorder. Cluster A personality subtypes are typically characterized by odd and eccentric behaviours. The Borderline, Antisocial, Histrionic, and Narcissistic Personality Disorders belong to the Cluster B Personality Disorders, which are characterized by erratic, dramatic and emotional behaviour. Finally, the Cluster C Personality Disorders are composed of the Avoidant, Dependent, and Obsessive-Compulsive Personality Disorders. The Cluster C personality disorders usually evidence anxious and fearful behaviours. For the diagnostic criteria of Borderline Personality Disorder, please refer to Appendix 1.

The diagnosis of Borderline Personality Disorder (BPD) has reasonable interrater and test-retest reliability (Davies & Akiskal, 1989), with moderate content validity for the BPD criteria set (Blais, Hilsenroth & Castlebury, 1997; see Appendix 1 for criteria set). Grueneich (1992) conducted a meta-analysis of 19 studies on borderline personality disorder and found high Kappa values, with a median value of 0.78 across studies. In general, the kappa, sensitivity, specificity, positive predictive power, and negative predictive power values for individual Borderline Personality Disorder symptoms were within the acceptable range.

The choice behind selecting this particular case of Borderline Personality Disorder involved great consideration. First, it was a complex, difficult case that was amenable to being convincingly portrayed by the researcher. However, the general diagnosis of Borderline Personality Disorder has a certain controversy surrounding it. Some have even referred to it as being a garbage can category. It must be said, however, that we wanted a case that was not clear-cut, that was difficult and challenging. A case that had several presenting problems that did not necessarily load directly onto one diagnosis. Given all the information available on this patient, there was one parsimonious diagnosis, and that was Borderline.

In addition to the clinical information the consulting psychiatrist provided, there were three actual tests on file that the psychologist administered to the patient when

the patient first came to see the psychologist. These were the WAIS, MMPI and the Rorschach. These tests were presented to subjects as part of the protocol since these were the tests the psychologist deemed worthy of administering at the time. Please refer to Appendix 4 for a copy of the stimulus material.

3.1.6 Determination and Conceptualization of Accuracy

In North America, the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; APA, 1994) is the main guide for mental health professionals to aid in the diagnosis of individuals (Dunne & Chute, 1999). The DSM conceptualizes mental disorders as being best represented by five axes of functioning, each one contributing orthogonal information as to the clinical picture of a patient. The use of a multiaxial system "facilitates comprehensive and systematic evaluation with attention to the various mental disorders [*Axis I and Axis II*] and general medical conditions [*Axis III*], psychosocial and environmental problems [*Axis IV*], and level of functioning [*Axis V*] that might be overlooked if the focus were on assessing a single presenting problem" (APA, 1994, p.25, *italics added*). Please refer to Appendix 3 to view the Multiaxial Evaluation Report Form.

For diagnoses derived from the DSM, Axis I and Axis II are the main Axes to consider with mental disorders. Axis I reflects clinical disorders, or other conditions that may be a focus of clinical attention such as Depression, Schizophrenia, Social Phobia and Alcohol Abuse. Axis II is for reporting personality disorders and mental retardation. The clinical presentation of the patient in this study was an Axis II personality disorder called Borderline Personality Disorder, previously explicated.

Empirical evidence suggests that the best manner to assess patients is to gather clinical information from multiple sources, using multiple methods and clinical instruments that are reliable and well validated (Meyer, Finn, Eyde et. al., 2001). For the purposes of developing a criterion diagnosis and cross validating the primary diagnostic portrayal of Borderline Personality Disorder, the simulated

patient (portrayed by the researcher) was assessed by independent clinicians (those not participating in the study). The simulated patient underwent several diagnostic evaluations administered by clinicians trained in using the most current and empirically validated diagnostic instruments and interviews available at the time, in order to determine what diagnoses the simulated patient would receive from each. These diagnoses were derived from the following instruments: the Diagnostic Interview Schedule (DIS), the Structured Clinical Interview for DSM-IV Axis I Clinical Disorders (SCID Interview), Self Administered Computerized SCID, the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II Interview), and the Millon Clinical Multiaxial Inventory-III (MCMI-III). The DIS, SCID, and SCID-II are structured clinical interviews administered by trained clinicians. The self-administered SCID is a computerized version of the clinician-administered SCID, which the patient completes on the computer, and the computer combines the clinical information and suggests diagnoses. The MCMI-III is a paper and pencil personality inventory consisting of true-false statements to be answered by the patient. The MCMI-III was specifically designed to measure the presence and severity of Axis II personality disorders. Finally, the diagnosis generated by the consulting psychologist, who treated the actual patient, was compared to diagnoses generated by these recommended clinical instruments. Thus, by obtaining confirmatory external diagnoses, this decreases the clinical uncertainty as to the diagnostic picture of the simulated patient and provides evidence that the simulated patient is being reliably portrayed by the researcher.

Diagnostic impressions of the study's participants were extracted from the think aloud protocols articulated during the diagnostic formulation section as well as from each subjects' written Multiaxial assessment. The diagnoses provided by subjects using the think aloud and multiaxial form methods were compared using three levels of diagnostic accuracy, to be described below.

Based on the findings of the validated clinical instruments, diagnostic accuracy was conceptualized as having three hierarchical degrees of specificity. The rationale for conceptualizing diagnostic accuracy using three cutoff points was to mirror the

University. The clinical experience of the participants in this group ranges from those who have completed their initial clinical training sequence, the four month practicum, equivalent to 600 hours of supervised clinical experience, to those who have recently completed their full year clinical internship, equivalent to about 2000 additional hours of supervised clinical experience.

Professional Group: (N=14; age = 50.4 ± 9.0 years; age range = 38 to 65 years). This group consisted of licensed clinical psychologists with an accumulation of at least ten years (postgraduate) full time clinical experience. For some analyses investigating extensive experience, this group was further subdivided into two groups based on the number of years practicing. The two groups were **Professional 10 years** (N=7; age = 44.4 ± 6.8 years; age range = 41 to 58; experience = 11.9 ± 2.7 years; experience range 9 to 16 years) and **Professional 25 years** (N=7; age = 56.4 ± 6.6 years; age range = 49 to 65 years; experience = 27.9 ± 4.8 years; experience range = 23 to 35 years).

3.2.2 Recruitment

Subjects for the Beginner group were recruited from a pool of McGill University undergraduates participating in research studies for course credit. Individuals interested in this particular study left their contact information on a sign-up sheet. Subjects for the In-training groups were recruited at two Montreal anglophone universities, McGill and Concordia. Finally, in order to obtain a representative sampling of experienced licensed Clinical Psychologists in the Montreal area, two Professional psychologist directories were cross-referenced: the Canadian Psychology Association (CPA) Directory, and the francophone "Ordre des Psychologues du Quebec Repertoire". Criteria for inclusion in the Professional group were as follows:

1) Holds Ph.D. or Psy.D. in clinical psychology
2) Accumulation of at least 10 years full time clinical experience
3) Specializes in adult psychopathology

Recruitment of subjects entailed contacting them via email or telephone, outlining the experiment, and the estimated time requirement (about two hours). Once a subject agreed to participate, a date and time were set. Subjects were tested either in their office or laboratory, or in the researcher's laboratory, whichever was most convenient to the subject. Before commencing the experiment, signed consent was obtained, according to IRB requirements.

3.2.3 Materials

The Problem Case. A senior and well-respected clinical psychologist in the Montreal area, with both university and hospital affiliations, was consulted to aid in the development of a clinical case. After much deliberation, the problem case was chosen and was based on an actual patient whom the consulting psychologist assessed and treated for a number of years. The problem case used in this study was the portrayal of a female in her twenties meeting diagnostic criteria for Borderline Personality Disorder. Please refer to Appendix 1 for the list of diagnostic criteria for Borderline Personality Disorder. The clinical details of the problem case can be found in Table 3.1. Content areas correspond to coding categories (see Section 3.3, Data Analysis for full explanation of coding categories).

Table 3.1 Description of the Problem Case – Page 1

CONTENT AREA	DETAILS of the PROBLEM CASE
Presenting Complaints: History of Illness	"Tara" is a 28-year-old woman, a pediatric nurse, who presents with stuttering, marital problems and symptoms of nervousness while doing presentations. She seeks help with her stuttering, which has generalized to all aspects of her life. She believes that her husband is having an affair, and he does not care for her anymore. They argue constantly about how to discipline their seven-year-old-daughter.
Previous Psychiatric History	She suffered from Anorexia Nervosa from the ages of 13 to 17, and was hospitalized at the age of 17 for three months for intensive treatment, because her weight was a mere 69 pounds.
Social History: Childhood & Growing Up	She was raised by a single mother, and never knew her father. She has no siblings. She suffered an incident of sexual molestation when she was 13, with a boarder in her mom's home. When she told her mother about the incident, the mom let him stay because "they needed the money". She fought with her mother all the time, but now realizes that much of it was due to her mother's illness (paranoid schizophrenia). She describes coming home from school and doing the groceries and cleaning the house because her mother never did those chores, she just sat around all day, with male friends coming over and the mom would smoke a lot. They were on welfare, and a social worker would come and visit them every so often to make sure that all was all right.
Educational History	Did well in school, and received her nursing degree via a CEGEP certificate

Table 3.1 CONTINUED Description of the Problem Case – Page 2

Current Living Situation	Lives with her husband of eight years and her seven-year-old daughter. On speaking terms with her mom, who sometimes baby-sits her daughter.
Work History	Worked at a local children's hospital for past 6 years. She is currently in the Haem-Oncology ward, specializing in immuno-compromised children. Before that, she worked as a translator (French/English).
Medical history	Hospitalized for Anorexia Nervosa and nearly died. No other medical problems besides childhood chicken pox, and the occasional cold/flu.
Family History: (Psychiatric, Medical)	Her mother was diagnosed as having paranoid schizophrenia when the mom was in her twenties. She did not know much about any other relatives and she did not know who her father was. The mother moved to Canada from Greece and is not in contact with her family back home.
Substance Abuse: drug & alcohol	None. Denies any alcohol or drug use/abuse past or present
Interpersonal Relationships	She has few, if any friends. She has instances where she becomes uncontrollably aggressive towards her husband. During these fits, she often loses recollection of the events, and she describes herself as "out of control".

Table 3.1 CONTINUED Description of the Problem Case – Page 3

Suicide Attempts	"Tara" tried to commit suicide once during her marriage. She had a fight with her husband several years ago when they were on vacation, so she stormed out of the house, decided she was going to swim out into the ocean and drown. She swam for a while, and thought she felt a fish touch her foot and this scared her so she swam back to shore. She describes being upset that her husband did not run after her to save her.
Mental Status	"Tara" feels alone, and empty inside, and may evidence signs of mild depression. She can become confrontational during interview and projects a sense of superiority and entitlement.
Disorder Specific	She sometimes engages in self-mutilating behaviour, such as slapping her own face. She sometimes binges and purges herself, about once or twice a year.

3.2.4 Procedure

The study was divided into two main parts, which used different tools for data collection and analysis. The first part of the study focused on collecting clinical knowledge measures and background information using checklists filled out by subjects and specific questions asked by the researcher. The second part of the study was the assessment task, which was audiotaped in its entirety.

Part 1 - Clinical Knowledge Measures & Background Information. The first part of the study focused on collecting variables believed to have an impact on clinical assessment behaviour. Subjects completed measures assessing their familiarity with psychometric instruments commonly used in psychological testing and assessment, and their experiences with various populations, clinical groups,

1) referral note (think aloud)
2) interview
3) request tests and justify (think aloud)
4) interpret tests (think aloud)
5) diagnostic formulation (think aloud)

The components of the assessment task can be found in Table 3.2, along with a detailed ccount of the verbatim instructions given to subjects at the beginning of each component. Each component corresponds to the sequential ordering of clinical activities that psychologists typically follow in the assessment of a new patient. In brief, the main components are referral note, interview, testing, test interpretation, and diagnosis. The last two components were added to gather more information regarding the assessment process, namely the confidence of the subject, and whether diagnosis differs when using the Multiaxial Evaluation Report Form (APA, 1994). The rationale behind adding an additional diagnostic extraction measure is that the Multiaxial Form prompts subjects to consider Axis II personality disorders. It is the Axis II personality disorders (Borderline Personality Disorder more specifically) which are considered to be accurate diagnoses for the purposes of this study. Further, the use of the Multiaxial Evaluation Report Form is recommended by the American Psychiatric Association as a useful aid when formulating a diagnosis (APA, 1994).

Table 3.2 Detailed Components of the Assessment Task

1- Referral note – Subjects were presented with the referral note and told "This is the referral note." The note contained the following information:

Name: Tara

Age: 28

Physical Description: attractive, well-groomed, average height, average weight

Born: Montreal, Canada

Language: Bilingual, French and English

Occupation: Pediatric Nurse

Complaint(s): stutters; anxious when giving presentations; marital discord

"Diagnostic work-up requested"

2- Interactive interview - where the researcher plays the role of the patient (verbally and non-verbally), and the subject, who understands that this is a simulation, is asked to behave as they normally would in their practice (or, in the case of a beginner, to behave as a psychologist would).

3- Request tests and justify – Subjects were asked "What tests or measures, if any, would you like to give the patient?" The subject is given the opportunity to indicate which tests, measures, or other information he/she would like to obtain, and to explain why

4- Interpret tests – Subjects were told "These are the actual test results from the patient whom this case is based on. They are the WAIS, with only full scale IQ available, the Rorschach, without inquiry, and the MMPI profile. Please use these tests as you would if they were given to you as part of the referral package." The three test results obtained from the patient file were presented to the subject to interpret (see Appendix 4 for the test material).

5- Diagnostic Formulation – Subjects were asked, "At this point, I would like to ask you what is your diagnostic formulation of the patient?" The subject was asked to provide (orally) a diagnostic formulation for the patient

6- Confid ce Ratings - subjects were asked to respond to the statement "How confident are you of your diagnostic impressions?" They were then given a seven-point Likert rating scale to indicate the level of confidence in their decisions, with 1 being very unconfident and 7 being very confident

7- DSM evaluation – Subjects were then presented with the Multiaxial Evaluation Report Form from the DSM IV (APA, 1994), and asked to fill it out

3.3 Methods of Data Analysis

The assessment task, as previously explicated, was audiotaped in its entirety, resulting in about 90 minutes of recorded material for each subject. These audiotaped sessions were transcribed verbatim for each subject.

3.3.1 Coding of the Interview Component of the Assessment Task

Transcriptions of the interview component of the assessment task were subjected to analysis using two coding schemes. The first focuses on the structure of the interview, in terms of types of questions asked by subjects, and the second on the content topics covered during the interview.

Coding of Interview Structure – A coding scheme was developed to investigate differences in the types of questions asked and statements made by subjects during the clinical interview. It is important to consider this structural information because it captures the direction of information flow. Questions typically are posed to *acquire* information while statements made by subjects are an attempt to *convey* information (factual or emotional) to the patient. Further, the type of questions

asked can influence the quality and quantity of verbal material elicited. Statements provided by subjects, as conceptualized in this study, can be of an informational or encouraging nature, either of which can impact the alliance between clinician and patient. A strong alliance enables a patient to feel comfortable about taking about their deepest thoughts and feelings.

This study used five categories for the coding of the interview structure, two for coding the type of questions, and three for coding the type of statements.

Question types are based on the definitions used in Patel, Evans, Kaufman (1989), where the authors present a cognitive framework for the examination of the doctor-patient interaction in medicine. For the present investigation, questions are coded as either open-ended or close-ended. The open-ended questions in this study roughly correspond to the 'Wh'-questions in the Patel, Evans & Kaufman framework, while the close-ended questions in this study roughly correspond to the 'Yes/No'-questions in the Patel, Evans & Kaufman framework. Essentially, open-ended questions, as defined in this study, are questions that elicit responses that are not one or two word responses, but invite elaboration or explanation from the person who the question is directed at. For example, questions such as "Tell me about your marriage", "How did you come to be hospitalized?" and "Why do you feel angry all the time?" would all be coded as open-ended questions in this study. Close-ended questions, as defined in this study, are questions that elicit one or two word responses, and by their nature, limit opportunity for elaboration. However, it may be that the patient chose to answer a close-ended question with elaboration. In either case, the nature of the question, rather than the response, formed the basis for coding. For example, "How old are you?", "Did you overdose on sleeping pills?" and "Can you swim?" would all be coded as close-ended questions, regardless of whether the response to these questions was brief (one or two words) or elaborate.

Statement types were coded using three categories. The first category, statements, was comprised of non-question utterances made by subjects with the aim of conveying information, or making a comment. For example, "I believe you are good mother" and "It seems that you have always had this pattern of intense relationships" would be coded as statements because they convey information to the

patient. The second coding category, encouragements, were non-question utterances with the aim of facilitating information flow from patient to the subject, as well as providing some sort of emotional support. Examples of encouraging statements include "That must have been very difficult for you" and "Uh-huh" or "Please continue". Finally, the las coding category was developed to code all other statements that do not fit into any other coding category for the interview structure. Typically, this "other" category was comprised of statements such as greetings or reiterations of what the patient already said.

Thus, the five categories developed for coding the interview structure were open-ended and close-ended questions, statements which conveyed information, statements of an encouraging nature, and other statements which did not fit into the current coding scheme.

Coding of Interview Contents - A coding scheme was developed to investigate differences in the kinds of clinical information acquired by subjects during the clinical interview. It is important to investigate the information sought by subjects during the interview, as it is this information (along with testing material in some circumstances) that forms the basis of clinical impressions and subsequent diagnostic considerations.

The coding scheme for examining the interview contents was based on literature used to teach interviewing skills to graduate students in clinical psychology that outlined what topics to cover during a comprehensive clinical interview of a mental health patient (Turner & Hersen, 1985). Morrison (1995) detailed how to conduct the first interview with a mental health patient, including which categories to cover and outlined specific material to be covered in order to formulate an official diagnosis. These proposed interview topics were used to code for content in the clinical interview for this study. Please refer to Table 3.3 for a complete list of coding categories used to code for information content in subjects' clinical interviews with the simulated patient.

Table 3.3 Coding Categories for the Information Content of the Interview – Page 1

CODE LABEL	CONTENT	EXAMPLE
A	Presenting Complaints: History of Illness	"Tell me about what brings you here today" "How long have you been stuttering?"
B	Previous Psychiatric History	"Have you ever taken antidepressants?" "What do you remember about being hospitalized for Anorexia?"
C	Social History: Childhood, Growing Up	"Tell me about your childhood" "Did you remember your grandfather?"
D	Educational History	"Did you do well in school?" "Where did you complete your nursing degree?"
E	Current Living Situation	"Do you own your own home?" "Who else lives with you?"
F	Dating/Marital History	"How many boyfriends did you have?" "What did fight about with your last love?"
G	Work History	"How many jobs have you been fired from?" "Tell me about your responsibilities at work"
H	Leisure Activities	"What do you do for fun?"
I	Legal History	"Have you consulted a lawyer yet?" "Were you ever in trouble with the law?"
J	Medical history	"Have you ever had surgery before?" "Do you currently take medication?"

of test data into diagnostic reasoning is recommended as a valid additional source of information for clinical psychologists. Thus, whether subjects sought testing material, the number of tests requested, what these tests were, and whether they referred to the testing material presented to them for interpretation while they formulated diagnoses was investigated.

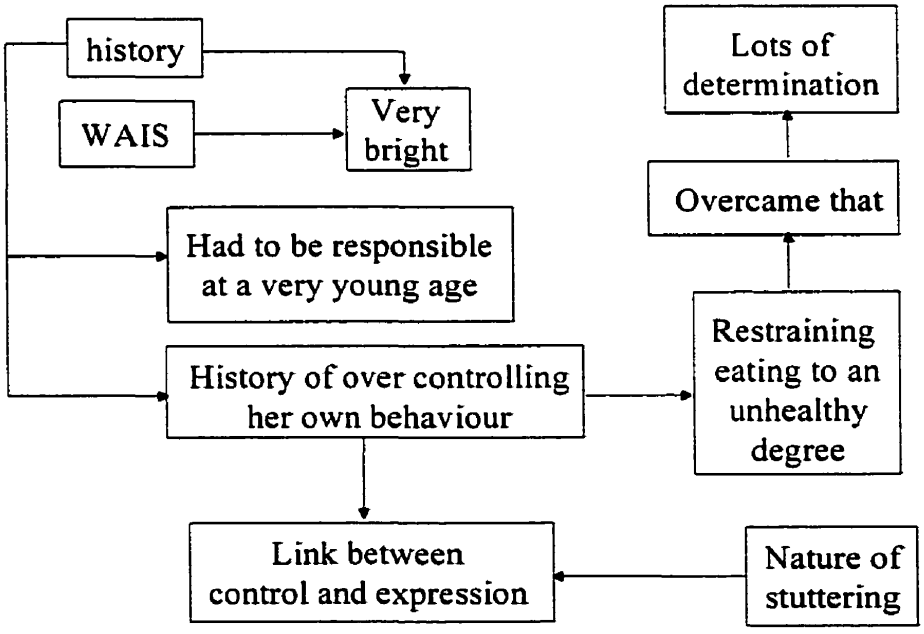
During the analysis of testing material, an interesting pattern in responding emerged. After the interview component of the assessment task, subjects were asked "What tests or measures, if any, would you want to administer to this patient?" Subjects responded by naming tests or indicating what additional information would be useful to have. In many instances, however, the answer to this question was embedded in responses where subjects outlined their diagnostic impressions, although diagnostic impressions were not asked for at this point. Thus, whether or not subjects provided detailed diagnostic impressions during the "Request tests" component was recorded. Subjects were then provided with three of the patient's test results to interpret, the WAIS, the MMPI, and the Rorschach. Remember that only these instruments were provided to subjects as only these instruments were administered to the actual patient by her treating psychologist. Subjects were then asked to provide a diagnostic formulation of the patient. Diagnoses provided during the "Diagnostic Formulation" component of the assessment task were compared to diagnoses provided during the "Request Tests" component to determine whether there was an increase in diagnostic accuracy. The increase in accuracy, if any, could be attributed to information gained with the introduction of testing material.

3.3.3 Development and Coding of Diagnostic Networks

As described above, after conducting the clinical interview, and reviewing psychological test material, subjects provided a diagnostic formulation of the patient prompted by the question "At this point, I would like to ask you what is your diagnostic formulation of the patient?" Each subject's response was recorded, transcribed verbatim, and a diagnostic network was generated. A diagnostic network is a type of data structure used to represent diagnostic

Figure 1 Example of Developing a Network from a parsed Diagnostic Formulation Excerpt

Hmmmm, well, from her history/ and her [WAIS]/, she is very bright/, had to be very responsible at a very young age/, a history of maybe over controlling her own behaviour/, in terms of restraining the eating to unhealthy degree/ then being able to overcome that/, that shows a lot of determination/. Perhaps there is some link between control and expression/, perhaps the nature of the stuttering is related to that/.



3.3.4 Statistical Considerations

Prior to analysis, all variables were screened for accuracy of data entry and missing values. Missing values were deleted from the analyses, an option endorsed by Tabachnick and Fidell (1996). The continuous variables were examined for violations of the ANOVA assumptions of homogeneity of variance using Levene's test for homogeneity of variance. To reduce extreme positive skewness, some variables were transformed. In the very few cases where data were found to have violations of homogeneity of variance assumptions, data were subjected to an appropriate nonparametric test to confirm results of ANOVA analyses. ANOVA results will be reported. Finally, categorical variables were analyzed using the nonparametric χ^2 statistic. For purposes of brevity and clarity, significant findings along with appropriate post-hoc analyses will be summarized in table format. In addition, trends towards significance will be included.

A final note about planned comparisons and type I error rates. Due to the relative small sample size (35 subjects), power analyses determined that this study is designed to detect only very large effect sizes, and cannot detect more subtle differences between groups. Therefore a Bonferonni correction or similar method of reducing the occurrence of type I error was not incorporated into the analyses. We wanted this study to open the door to future research into this area, and invite replication of these findings. Therefore, we wish not to limit the number of potentially significant findings by being overly cautious with type I error rates. It is also for this reason that trends towards significance will be reported, so that future studies with a more precise focus and greater number of subjects can potentially replicate some of the specific findings of this thesis.

3.3.5 Interrater Reliability

Interrater reliability (% agreement) was assessed for the coding of interview contents and for the construction of diagnostic networks. The raters were two graduate students in clinical psychology. One of the raters was the researcher (L.Z.), with experience in interview and network coding methods. The inexperienced rater was explained the coding scheme for interview contents and instructed in the

construction of diagnostic networks. The inexperienced rater was given a practice session where the nature of the discrepancies between the experienced rater and the inexperienced rater were examined. Attempts were made to arrive at an agreement about discrepancies between the results of the two coding schemes. Roughly, ten percent of the interviews and ten percent of the networks were randomly selected for coding for interrater agreement.

The criterion for agreement for interview contents was identical codes for the content of a given section of discourse. The number of discourse sections containing identical coding divided by the total number of discourse sections resulted in the % agreement for coding of interview contents. The criterion for agreement of networks was a node and direction of link unit. Thus, the number of identical nodes coupled with identical arrow directions was divided by the total number of nodes and links in the network to arrive at the % agreement for the diagnostic networks. The % agreement between raters for the coding of interview contents was 73.82 % and the agreement between raters for the diagnostic networks was 69.56 %. This indicates a reasonable amount of agreement given the complex nature of this kind of data coding. The balance of the data was coded by the researcher/experienced coder.

CHAPTER IV

RESULTS AND DISCUSSION

The organization of this chapter follows the sequence of tasks presented to subjects in this study. First, in section 4.1, an examination of differences in the amount of training and clinical preparation is presented. These data were acquired during Part 1 of the study, where questionnaires and targeted questions were used to obtain this data.

The remainder of the results section presents and discusses findings generated from Part 2 of the study, the assessment task. In section 4.2, the time it took subjects to complete the components of the assessment task is investigated. The clinical interview with the simulated patient is characterized in section 4.3, along with excerpts of interview material of representative subjects from each level of experience. Section 4.4 presents results from the diagnostic formulation component of the assessment task. Diagnostic networks are examined and group differences discussed. Section 4.5 investigates the diagnostic accuracy of diagnoses provided by subjects during the think aloud diagnostic formulation section, as well as in written form on the Multiaxial Evaluation Report Form. Characteristics of participants that provided accurate diagnoses are discussed. Use of testing material is explored in section 4.6. Section 4.7 looks at subjective variables of subjects, such as their rating of the ecological validity of the experiment and whether they sought feedback about whether their diagnosis was correct. Section 4.8 presents a model of clinical assessment behaviour that captures in a realistic manner, the assessment process in psychology. The final section explores the role of extensive clinical experience in the assessment process and looks at differences between professional psychologists with ten years versus 25 years clinical experience.

Table 4.1 Descriptive and Clinical Training Variables – Significant Differences

Variable of Interest	Level of Experience			Group Differences
	Beginners	In-Training	Professional	
	x ± SD or (%)	x ± SD or (%)	x ± SD or (%)	
Age (years)	22.0 ± 4.60	29.77 ± 5.05	50.43 ± 8.95	F[2,32]=53.324, p=0.000 [a,b,c]
Years full time experience	0.00 ± 0.00	2.023 ± 1.15	19.86 ± 9.11	F[2,32]=42.761, p=0.000 [a,b,c]
Current # clinical hours/week	0.813 ± 1.19	4.462 ± 6.07	24.54 ± 13.6	F[2,32]=22.016, p=0.000 [b,c]
Test Familiarity Index	1.34 ± 0.371	4.07 ± 0.889	4.27 ± 1.18	F[2,32]=27.66, p=0.000 [a,b]
Specializes in Personality Disorders? (% Yes)	0.00	46.1	64.2	χ^2[2]=8.683, p=0.013 [a,b]

[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

Given that three groups were compared, all significant findings were subjected to post-hoc analyses, in particular, Student-Newman-Keuls post-hoc analyses, so that the nature of the significant findings (which groups differ from each other) could be explored. The Student-Newman-Keuls post-hoc analyses revealed the following

differences between groups on the measure of age. The Beginners were significantly younger than both the Professional (22.0 years versus 50.43 years) and the In-Training groups (22.0 years versus 29.77 years), and the In-Training group was significantly younger than the Professional group (29.77 years versus 50.43 years). This indicat that there is a significant age difference between all the groups. The nature of this difference is linear whereby less experienced groups tended to be younger and the more experienced group (the Professional group) tended to be older, as anticipated.

For the measure of number of years practicing full time, post-hoc analyses revealed that the Professional group evidenced significantly more experience (19.86 years) than the Beginners (0.00 years), or the In-Training group (2.02 years). This indicates that the Professional group is significantly more experienced than the two other groups studied. This finding was also anticipated as subjects were grouped according to level of experience in clinical psychology.

The number of hours of clinical work subjects engaged in per week was analyzed and it was found that the Professional group currently worked significantly more clinical hours (24.5 hours) than the Beginner (less than 1 hour) or In-Training (4.46 hours) groups. This indicates fundamentally different daily activities between the practicing Professionals and the In-Training group. Clinical psychology graduate students may spend parts of the year immersed in clinical work, treating and assessing patients on a full time basis. However, given that they are still pursuing graduate studies, this clinical work is not yet a continual part of their everyday activities.

Differences in familiarity with psychological tests indicated a significant difference between the Beginner group (1.34) and each of the other groups. The In-Training group (4.07) and the Professional group (4.27) did not differ significantly from each other on this measure of test familiarity. This singular difference between groups on the variable of test familiarity indicates that undergraduates, with zero training in clinical psychology, do not endorse any familiarity with psychological tests and instruments. The fact that the other groups do not differ on this measure reflects the similar endorsement of participants with clinical training to be equally

familiar with this particular sampling of psychological tests and instruments (see Appendix 2 for a list of tests used to derive this familiarity index). Further, more clinical experience does not equate with a greater degree of familiarity with testing materials.

Whether subjects specialized in Axis II personality disorders was important to consider since the target diagnosis was an Axis II personality disorder. Zero subjects in the Beginner group reported that they specialized in Axis II Personality Disorders, while 46.1% (6 out of 13) subjects reported specializing in Personality Disorders in the In-Training group. The majority of those in the Professional group (64.2% or 9 out of 14 subjects) reported specializing in Personality Disorders. This difference between groups on the variable of specialty in personality disorders indicates that undergraduates, with zero training in clinical psychology would not endorse any specialty in personality disorders. The fact that the other groups do not significantly differ on this measure reflects the similar endorsement of participants with clinical training to specialize in personality disorders. This non-difference between the clinically initiated groups allows unbiased analysis of diagnostic accuracy, as the primary diagnosis was a personality disorder.

4.2 Time to Complete Tasks

The time it took subjects to complete the components of the assessment task (referral note, interview with simulated patient, request tests, interpret tests, diagnostic formulation, and Multiaxial evaluation) were recorded to the nearest 30 seconds. There were no limits as to the amount of time a subject could spend on any component of the assessment task. Therefore, subjects indicated when they were finished.

In order to determine whether there were significant differences between groups on the measures of interest, Analysis of Variance procedures were performed. There was found to be significant differences between the Beginner, In-Training and Professional groups for the singular measure of Time for Diagnostic Formulation ($F[2,32]=5.510$, $p=0.009$). The ANOVA analyses failed to detect a significant

The lack of differences between groups on the remaining measures of time to complete the Referral Note, Interview, Request Tests, Interpret Tests, Certainty Rating, and Multiaxial Evaluation Form sections of the clinical assessment indicates no detected effect of clinical experience.

4.3 Characterization of the Interview

A clinical interview is an integral part of any psychological assessment. In a survey of practicing and teaching clinicians, the ability to conduct a comprehensive interview was ranked as the most important skill, out of 32 others, needed by mental health practitioners (Morrison, 1995). Great efforts were taken to make the interview process as realistic as possible. The measures extracted from the interview protocols were meant to capture the overall process of interviewing a patient. Important topics to investigate were 1) interview structure: the number of questions, the type of questions (open versus close ended questions), and type of statements (informational or encouraging), and 2) interview contents: those content topics deemed important to any comprehensive clinical interview in psychology or psychiatry.

4.3.1 Structure of the Interview

Each interview was coded for structure using the following five coding categories (please refer to section 3.3.1 for an explanation of the coding scheme) 1) open-ended questions, 2) close-ended questions, 3) statements made by the subjects where they provided information to the patient, 4) encouraging statements from subjects which encouraged or facilitated the flow of information from the patient to subject, and 5) other (not otherwise coded) statements. In order to control for the length of the interview, the raw counts of the five variables of interest were divided by the total number of utterances made by subjects during the interview, resulting in the "proportion" of the variables of interest. Thus, the proportion of open-ended and close-ended questions, the proportion of statements, the proportion of encouraging

statements and the proportion of other statements were subjected to statistical analyses to determine whether any group differences exist.

One-way ANOVAs were performed and there was found to be significant differences between groups across levels of experience for the variables of Proportion of Open-Ended Questions ($F[2,12]=7.004$, $p=0.010$), and Proportion of Close-Ended Questions ($F[2,12]=20.021$, $p=0.000$). There were trends towards significance noted for the following variables: Total Number of Utterances ($F[2,12]=3.553$, $p=0.061$), Proportion of Statements ($F[2,12]=3.837$, $p=0.051$), and Proportion Uncodable Statements ($F[2,12]=3.378$, $p=0.069$). The Analysis of Variance failed to detect a significant difference between the Beginner, In-Training and Professional groups for the variable of Proportion of Encouraging Remarks ($F[2,12]=0.238$, $p=0.792$). Please refer to Table 4.3 for a summary of significant findings as well as findings that demonstrate a trend towards significance.

Table 4.3 Summary of Significant Findings and Trends for the Interview Structure

Variable of Interest	Level of Experience			Group Differences
	Beginners	In-Training	Professional	
	x ± SD	x ± SD	x ± SD	
Proportion (Open-Ended Questions)	0.230± 0.060	0.437±0.077	0.341± 0.101	F[2,12]=7.004, p=0.010 [a,c]
Proportion (Close-Ended Questions)	0.465± 0.049	0.235± 0.047	0.296± 0.056	F[2,12]=20.021, p=0.000 [a,b]
Total # of Utterances	69.33± 32.13	154.7± 56.23	145.0± 42.10	F[2,12]=3.553, p=0.061 [a,b]
Proportion (Statements)	0.132± 0.066	0.147± 0.073	0.275± 0.114	F[2,12]=3.837, p=0.051
Proportion (Uncodable Statements)	0.080± 0.076	0.033± 0.024	0.014± 0.014	F[2,12]=3.378, p=0.069 [b]

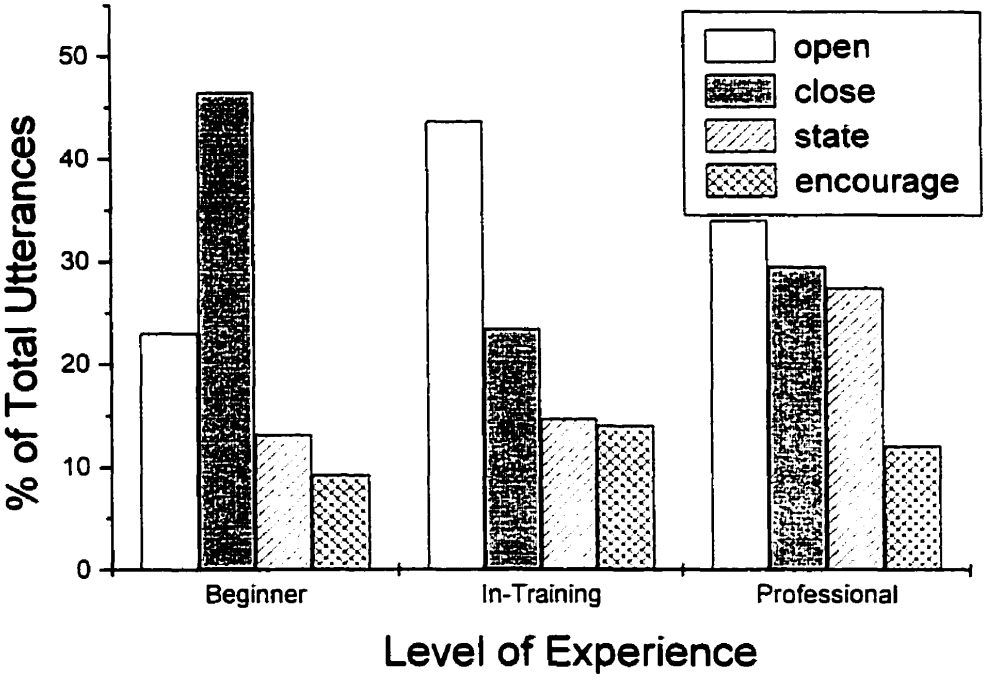
[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

ask. This could function to decrease the number of utterances, or verbal exchanges during the interview.

A trend towards significance was noted for Proportion of Statements ($F[2,12]=3.837$, $p=0.051$). This variable was subjected to Student-Newman-Keuls post-hoc analyses to examine group differences, and results indicated no significant differences between the Beginner group, the In-Training group and the Professional group on this measure. However, compared to the other groups, the Professional group had almost double the proportion of statements (27.5% versus 13.2% for Beginners and 14.7% for the In-Training group). Although it approaches significance, this finding indicates a tendency of Professionals to provide information to the patient.

These findings imply a possible generalized strategy with increasing clinical experience to roughly divide the amount of verbal exchanges across the four variables (open and close-ended questions, statements and encouraging statements). Please refer to Figure 2 (next page) for a graphical depiction of this trend.

Figure 2 Types of Questions and Statements employed during the Clinical Interview as a Function of Clinical Experience



Note: bars do not add up to exactly 100% as these numbers are the sums of group averages

In the above graph, the Beginners have a preponderance of close-ended questions, with the remainder of verbal exchanges (about 55% of total utterances) being dedicated to other types of information transfers. The In-Training group demonstrates a modified reversal of this trend, but still spends a large part of the verbal exchanges in asking open-ended questions, leaving about 55% of their information transfer opportunities for the other three types of verbal exchanges. Finally, the most experienced group appears to spread their verbal exchanges across all four kinds of information transfers more equally, spending most of their efforts using open ended questions (34.1%) but leaving the vast majority of their information transfer opportunities (roughly 69%) for the other three categories. Of course, the clinical interview is meant to be a source of information for the

diagnostician, and this is why all subjects spent the majority of all their verbal exchanges asking questions of the open or closed variety (Beginners = 87.5%, In-Training = 67.2%, Professionals = 63.7%), with a minority of verbal exchanges dedicated to statements of encouragement or providing the patient with information (Beginners = 22.5%, In-Training = 28.8%, Professionals = 39.6%).

To test whether the Professionals' interviews are characterized by a tendency to spread types of questions across all four variables (open, close-ended questions, statements, encouragement), the relative differences between each of the four categories was summed within each group (Beginner, In-Training and Professional) to yield a total relative difference. A one-way ANOVA was computed and there was no significant difference between groups on the measure of total relative difference ($F[2,12]=1.705$, $p=0.223$), indicating that there is no generalized tendency for Professionals to spread the type of questions and statements more evenly throughout the interview.

To determine whether there are differences across levels of experience in terms of type of questions asked (open-ended versus close-ended), paired samples t-tests were performed for each of the three groups. For the Beginners, one-tailed paired sample t-tests revealed significantly more close-ended questions were asked during the interview ($t[2]=-8.297$, $p=0.007$). The In-Training group evidenced the reverse: one-tailed paired sample t-tests revealed significantly more open-ended questions were asked during the interview ($t[5]=4.527$, $p=0.003$). Finally, one-tailed paired sample t-tests performed on the question data of the Professional group indicated that the Professionals did not evidence a bias or preference to asking more of either type of question ($t[5]=0.080$, $p=0.470$).

Finally, to determine whether there are differences across levels of experience in type of statements made (providing information to the patient (statements) versus encouraging statements) paired samples t-tests were performed for each of the three groups. For the Beginners, one-tailed paired sample t-tests revealed no significant difference between the number of statements and the number of encouraging remarks ($t[2]=0.429$, $p=0.355$). The same for the In-Training group: one-tailed paired sample t-tests revealed no significant difference between the number of

statements and the number of encouraging remarks (t[5]=0.108, p=0.459) asked during the interview. Finally, one-tailed paired sample t-tests performed on the statement data of the Professional group indicated that the Professionals provide significantly more information to the patient (statements) then encouraging remarks (t[5]=2.329, p=0.034). This indicates a relative preference or bias of the Professionals to provide information to the patient rather than encouraging remarks, whereas for the Beginner and In-Training groups, there were no significant differences between these two variables.

4.3.2 Interview Contents

The content of the interview, in other words, the type of information extracted from the patient was also of considerable importance, as the clinical data obtained from the interview forms the basis of diagnostic considerations. Therefore, the interviews were coded using a scheme based on recommended topics to cover during a comprehensive clinical interview in psychiatry or psychology. Section 3.3.1 (Methods Section) provides a detailed explanation of the coding categories, with illustrative examples of each. Once the interview protocols were coded, the number of utterances pertaining to a given coding category was tabulated. To be able to control for the length of the clinical interview, the raw counts of each coded category for a given subject were divided by the total number of utterances for that subject, resulting in "proportions" of the variables of interest. Finally, in order to obtain a general index of the extent to which all or most of the recommended topics were covered by subjects, a Completeness Index was developed. In short, the number of different topics covered by a given subject (even if they asked one question about the topic) was divided by the total number of topics, 18. For example, if a subject asked questions solely about Presenting Complaints, Childhood, and Suicide, then the Completeness Index would be 3/18 = 0.167, or 16.7 % of recommended topics were covered by this subject.

To examine group differences in the proportion of verbal exchanges containing each of the interview topics, Analysis of Variance procedures were used. ANOVA

detected significant differences between groups for the following variables: Leisure Time Activity questions (F[2,31]=5.443, p=0.009), Disorder Specific questions (F[2,31]=5.026, p=0.013), and Completeness Index (F[2,31]=10.392, p=0.000). There was a trend towards significance noted for Mental Status Exam questions (F[2,31]=2.837, p=0.074). The Analysis of Variance failed to detect a significant difference between the Beginner, In-Training and Professional groups for the remainder of the interview topics. Please refer to Table 4.4 (next page) for a summary of significant findings, along with group means.

Table 4.4 Summary of Significant Differences between groups in Interview Topics

Variable of Interest	Level of Experience			Group Differences
	Beginners	In-Training	Professional	
	Mean ± SD	Mean ± SD	Mean ± SD	
Leisure Time Activities	0.0034± 0.0049	0.0152± 0.020	0.00± 0.00	F[2,31]=5.443, p=0.009 [a,c]
Disorder Specific	0.0025± 0.007	0.0640± 0.052	0.0229± 0.050	F[2,31]=5.026, p=0.013 [a,c]
Completeness Index (1 to 18)	8.625± 2.56	13.33± 1.97	10.00± 2.72	F[2,31]=10.392, p=0.000 [a,c]

[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

The variable of Leisure Time Activities questions was subjected to post hoc analysis to determine which groups differed significantly from each other. Post-hoc analyses revealed there to be significant differences between the In-Training group and the other groups. The Professional group did not ask about leisure time

comprehensive clinical interview. This is evidenced by their significantly higher Completeness Index. Further, in order to increase the reliability of psychiatric diagnoses, questions regarding diagnostic criteria are highly recommended to enable a diagnostician to formulate an accurate and reliable DSM-IV diagnosis. Perhaps an awareness of the literat e on increasing diagnostic reliability and the subsequent emphasis on training interview strategies in the graduate programs of the universities sampled in this study permits subjects from the In-Training group to conduct comprehensive interviews, with disorder-specific questioning. These factors can influence diagnostic accuracy, which will be addressed in section 4.5.

4.3.3 Illustrative Interview Portions

Excerpts of representative verbal protocols are used to illustrate stylistic and strategic differences across levels of clinical experience and to explore characteristic variations in interviewing not easily amenable to quantitative measurements. The various comments are found in italics next to subjects' interview questions. The following excerpts are taken from the initial minutes of the clinical interviews. Discussions about the excerpts follow. Please refer to Figure 3 for the initial moments of an interview with a Professional subject, to Figure 4 for the initial moments of an interview with a Beginner subject, and Figure 5 for the initial moments of an interview with an In-Training subject.

Figure 4 Transcript of a Beginner Subject Commencing the Interview

S: Ok, hi my name's X.	*Introduction, followed by pause, so patient responded with greeting*
T: Hi.	
S: Um, ok I'd just like to ask you first of all some questions about your marriage, with your husband.	*Preamble about what is about to be asked (more likely putting herself at ease, than an attempt at putting the patient at ease)*
T: Uh hum.	
S: Um, how long have you been married for?	
T: For eight years.	
S: For eight years, and how's your marriage been for the past eight years? Would you say that it's a stable relationship?	*Asks open ended question, then follows it up with a close ended question*
T: Um, uh at the beginning things were much better. We were happier together.....Um he seemed to.... to .. he seemed to listen to me more	
S: Uh hum.	*Continues to listen*
T: ..And um ..now um I feel uh I feel very much alone in the marriage .. um and we fight constantly. We're always fighting and what we fight most about is disciplining ...how to discipline our daughter Christina.	
S: Ok, um .. How are you getting.. Ok actually..um .....how long have you been a pediatric nurse for?	*Overwhelmed by the sensitive information. No supportive statement. Abrupt change in topic*

Figure 5 Transcript of In-Training Subject Commencing the Interview

S: Hello Tara, my name is X, I've been working at X for a number of years, and I think we should probably get started now. Can you just tell me a little bit about yourself?	*Introduction, explanation of job security, followed by too vague a question*
T: What do you want to know?	
S: Oh, well, I guessv sgood thing to start with would be what brings you here today? Like what's, what's mainly on your mind?	*Narrows the question, and asks it two different ways*
T: What's, what's mostly bothering me now is that I'm stuttering.	
S: Uh-huh.	*Encourages patient to continue*
T: The stuttering began three months ago, while I was given extra responsibility at work to give, presentations to, to the incoming nurses.	
S: Uh-huh.	*Encourages patient to continue*
T: And, I welcomed this additional responsibility, but I found that about one hour just before my first presentation I felt a bit nervous and apprehensive, but I thought that was, that was fairly normal, because I'm not habituated to doing presentations, so I got up in front of the group.	
S: Uh-huh.	*Encourages patient to continue*
T: And, began my presentation, and out of my mouth, I started stuttering. Very unusual, because I had never stuttered before.	
S: So, that was actually the first time in your life you'd ever experienced a stutter...	*Reiterated what patient already stated*

gather much information, and while engaged in this process, they may get distracted by the details instead of the larger picture.

4.4 Characterization of Diagnostic Formulation

As described in the Methods section, the think aloud diagnostic formulation protocols were analyzed and netwo s were developed. From completed networks, several variables of interest were extracted. These were 1) the number of separate pieces contained in the diagnostic network, 2) whether the diagnostic network was a whole, interconnected piece, 3) the number of chunks contained in the diagnostic network (chunk = four or more interconnected nodes in the network) and 4) the number of "loose-ends" in the diagnostic network ("loose-end" = three or fewer interconnected nodes). The number of diagnoses contained in the diagnostic networks was also enumerated.

4.4.1 Diagnostic Network Structure

Statistical analyses failed to detect any significant differences between the Beginner, In-Training and Professional groups on the variables of interest: Total Number of Network Pieces ($F[2,30]=0.697$, $p=0.506$), Number of Loose Ends ($F[2,30]=0.392$, $p=0.679$), Number of Diagnoses Provided ($F[2,30]=1.476$, $p=0.245$). However, there were trends noted on the following variables: Whether the Networks were completely Connected (Whole) (χ^2 [2]=5.614, $p=0.060$), and Number of Chunks ($F[2,30]=2.992$, $p=0.065$). The trend of having a completely connected diagnostic network was more likely for the Professional group, while the trend of having the greatest number of chunks in the diagnostic network was most likely for the In-Training group. Please refer to Table 4.5 (below) for a summary of noteworthy trends and group means.

Table 4.5 Diagnostic Network Structure Descriptions-Trends

Variable of Interest	Level of Experience			Group Differences
	Beginners	In-Training	Professional	
	Mean ± SD (or %)	Mean ± SD (or %)	Mean ± SD (or %)	
Completely Connected Network (% Yes)	25 % (2 out of 8)	0 % (0 out of 12)	38.4 % (5 out of 13)	$\chi^2[2]=5.614$, p=0.060 [c]
# Chunks (> 3 nodes)	1.38± 1.06	2.50± 1.09	1.38± 1.50	F[2,30]=2.992, p=0.065

[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

To briefly summarize, a surface analysis of the diagnostic networks entailed an examination of differences in network structure. Differences in network structure are related to levels of comprehension and degree of cohesive integration of material into each subject's knowledge base. More of the Professionals (38.4%) than Beginners (25%) or In-Training group (0.0%) provided a completely integrated network ($\chi^2[2]=5.614$, p=0.060) indicating a trend for the Professionals toward cohesion of diagnostic concepts. Post-hoc analyses revealed that significantly more of the Professionals compared to the In-Training group generated a completely connected network. There were no significant differences on post-hoc analyses between the Professionals and the Beginners on this measure of cohesive diagnostic networks. These findings support the notion that Professionals, with their greater degree of clinical experience, demonstrate cohesion of their clinical knowledge about the patient during diagnostic considerations when compared to their lesser experienced counterparts, the In-Training group. The fact that Beginners and the Professionals perform similarly is consistent with findings comparing novices,

intermediates and experts in the medical domain (Schmidt & Boshuizen, 1993). This phenomenon, also known as the "intermediate effect", describes the development of medical expertise as a progression through a series of phases, and that the road to becoming an expert physician is not linear for all aspects of learning (Schmidt & Boshuizen, 1993). This progression entails the incorporation of medical knowledge into memory stores, but the organization of this newly acquired knowledge is not well developed. Thus, although information stored in the memory of intermediates and experts is of relatively similar *content*, the organization of the information is vastly more systematized and constituted in the experts whose years of experience helped to structure and enrich this knowledge over time. This can explain why, in this thesis, the Beginners and the Professionals perform similarly on a measure of clinical information organization, while the Intermediates present with a less organized knowledge base of the case.

The final measure was the number of chunks in the diagnostic networks. Although a trend was detected for this measure across the three groups ($F[2,30]=2.992$, $p=0.065$), post hoc analyses did not detect significant differences between pairings of the three groups.

4.4.2 Qualitative Examination of Representative Diagnostic Networks

The diagnostic networks were generated from each subjects' respective verbal protocols relating to the diagnostic formulation. The diagnostic formulation was collected only after the subject (in the following order) viewed the referral note, conducted an interview with the simulated patient, and viewed test material. Figures 6, 7 and 8 illustrate representative networks of subjects from each of the clinical experience groupings: Beginner, In-Training, and Professional. It should be noted that the boxes represent nodes, or clinical concepts, while the arrows represent the relations among the clinical concepts. A special category of node, as indicated by the oval frame, represents a diagnostic concept.

Figure 6 Diagnostic Network of an In-Training Subject

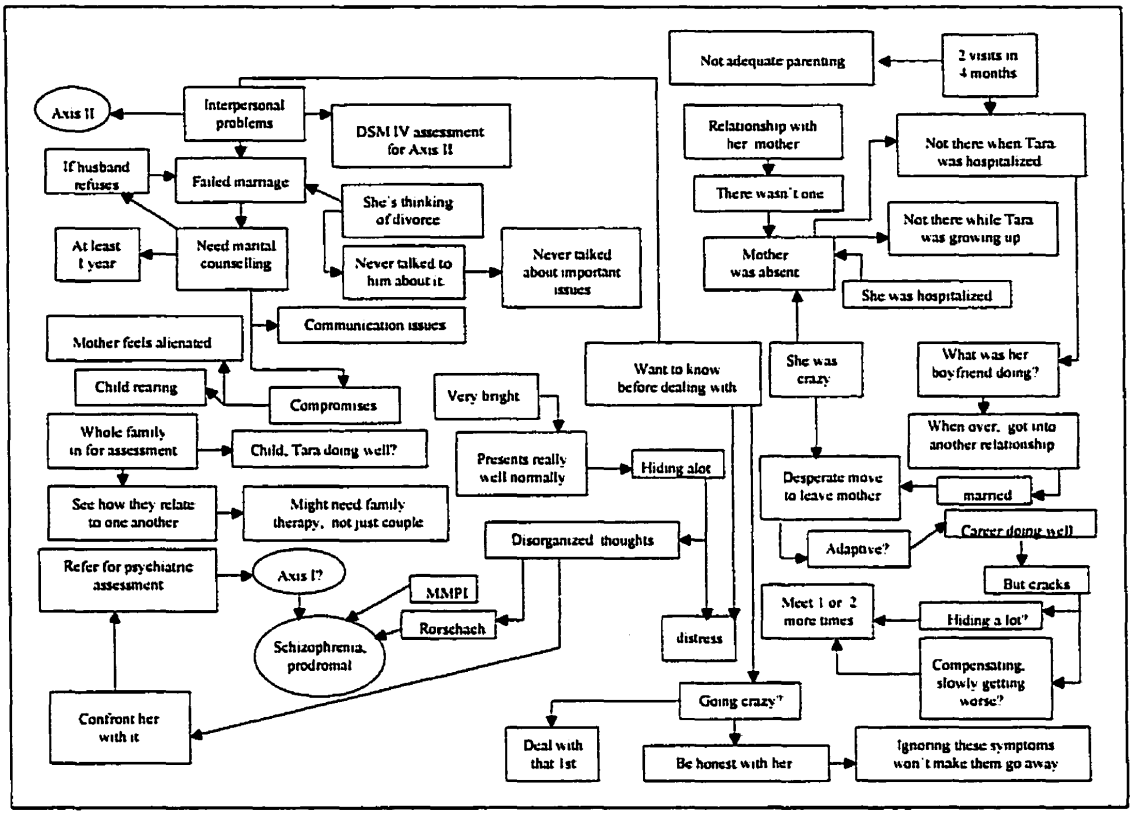


Figure 7 Diagnostic Network of a Professional Subject (below)

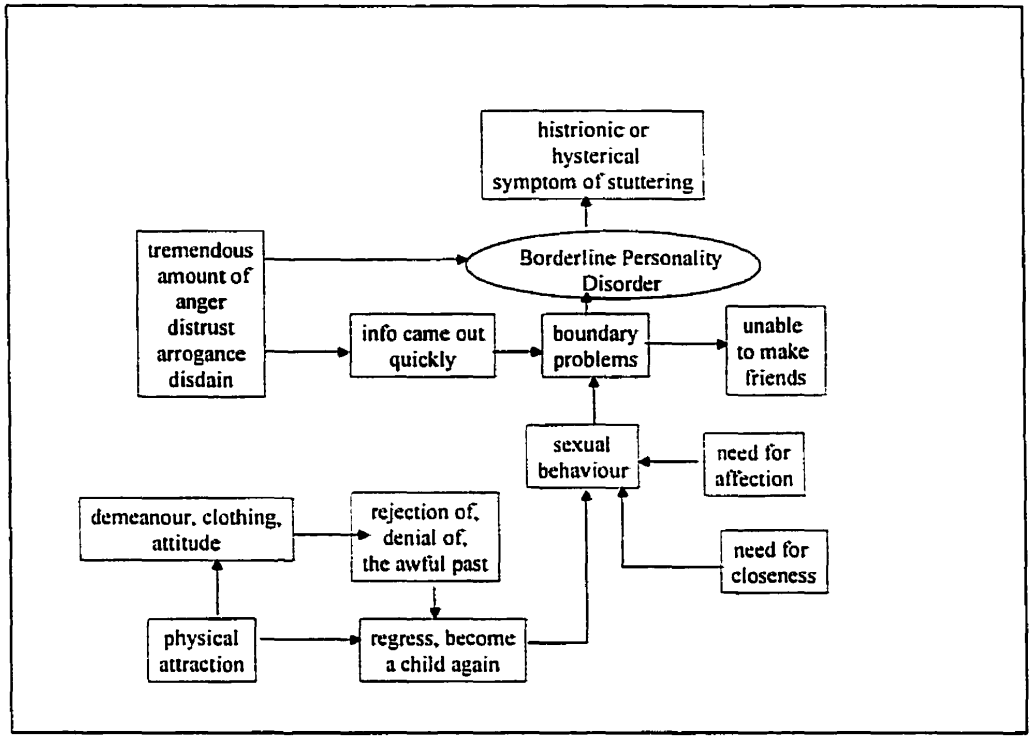


Figure 8 Diagnostic Network of a Beginner Subject (below)

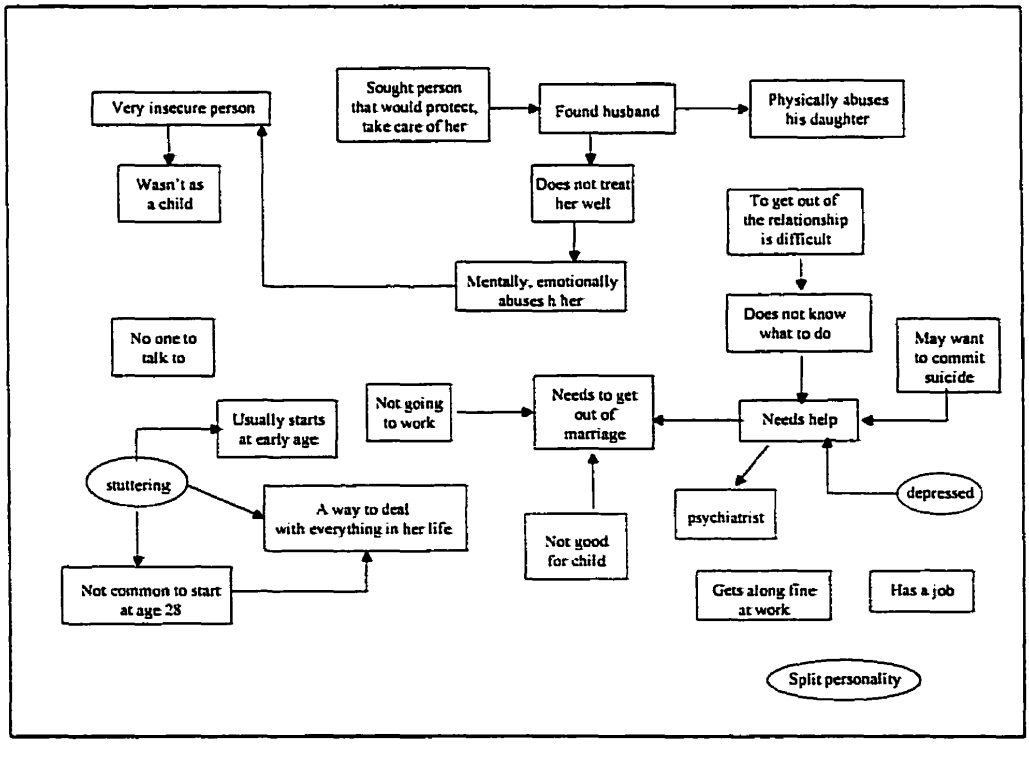


Figure 8 illustrates a diagnostic network of a typical Beginner, someone with no clinical experience. In this network, one can see that there are several "loose-ends" (3 or fewer interconnected nodes). The presence of loose ends indicates that the concepts contained in them are not yet fully integrated. One important finding in this particular network, is the fact that the subject made an incorrect inference as to the patient's husband physically abusing his daughter (upper right corner). This concept was not part of the information provided to the subject, but the subject came this inaccurate conclusion based on related information.

In sum, this qualitative analysis found that Beginners tend to provide diagnostic networks with loose ends, and may make incorrect inferences. The In-Training subjects may provide whole, interconnected networks, but the contents of the nodes indicate a path to obtain more information, with several uncertainties. The Professionals tend to also provide whole, interconnected networks, but there are fewer nodes. The nodes contain information as to personality dynamics, which support diagnostic notions. Although these findings are generated from single subjects, they can be used to support quantitative findings, and they can be a rich source of hypotheses to be tested in future studies.

4.5 Determination of Diagnostic Accuracy

As explained in the method sections 3.1.5 and 3.1.6, the clinical presentation of the simulated patient in this study was an Axis II personality disorder called Borderline Personality Disorder. Briefly, the Axis II personality disorders are grouped into three clusters based on behavioral similarities. The Borderline Personality Disorder falls into Cluster B typically characterized by erratic, dramatic and emotional patterns of behaviours. For the purposes of cross validating this primary diagnostic portrayal of Borderline Personality Disorder, the researcher compared diagnoses the simulated patient received from empirically validated diagnostic instruments and interviews. These instrument-derived diagnoses, along with the diagnoses provided by the consulting psychologist who treated the actual patient, can be found below in Table 4.6.

Table 4.6 Definitive Diagnoses and Possible Diagnoses generated by three Clinical Instruments designed to capture Axis I Disorders and two Clinical Instruments designed to capture Axis II Personality Disorders, compared to Diagnoses generated by the Consulting Psychologist

Source	Definitive Diagnoses	Possible Diagnoses to Explore (Rule Out)
DIS Interview	Anorexia - past	Depression Antisocial PD
SCID Interview*	Anorexia - past	Dysthymia, Social Phobia, Anxiety Disorder NOS, Bulimia
SCID Computerized	Anorexia - past	Depression, Bulimia Antisocial PD, Transsexual
SCID II Interview**	Borderline PD	Narcissistic PD
MCMI-III profile	Borderline PD Narcissistic PD	None
Consulting Psychologist	Borderline PD Anorexia – past	None

Note: Axis I instruments: DIS interview, SCID interview, SCID Computerized
Axis II instruments: SCID II interview, MCMI-III profile

* SCID-I also revealed as definitive diagnosis Mixed Anxiety, Depressive Disorder, which is a diagnostic category being researched for inclusion in the next DSM. Further, the following clinically significant episodes were suggested as possible diagnostic considerations by the SCID-I: Binge Eating, Minor Depressive Episode, Manic episode – past, Hypomanic episode – past.

** SCID-II also revealed as definitive diagnosis Depressive Personality Disorder, which is a diagnostic category being researched for inclusion in the next DSM.

typically demonstrate test signs that reflect thought disorder, impulsivity, anger, suspiciousness, depression, anxiety, and disturbed object relationships (Gartner, Hurt, & Gartner, 1989). It is important to keep these clinical symptoms in mind when viewing the collection of diagnoses provided by subjects as some of the diagnoses are influenced by information apparent in the tests, but not obvious in the interview. For example, on interviewing the simulated patient, a subject might not get the impression that she was schizophrenic; however, upon viewing the MMPI profile and Rorschach responses, one might suspect there to be some degree of paranoia. Thus, in this case, although the correct diagnosis is not Schizophrenia, there is some evidence to include Schizophrenia as a provisional or rule-out diagnosis because it is consistent with some of the clinical data. For a detailed account of diagnoses provided by subjects, see Figures 9, 10 and 11. Figure 9 presents a summary of diagnoses provided by Beginner subjects, Figure 10 presents a summary of diagnoses provided by In-Training subjects, and Figure 11 presents a summary of diagnoses provided by Professional subjects.

Figure 9 Summary of Diagnoses provided by Beginner Subjects (N=8)

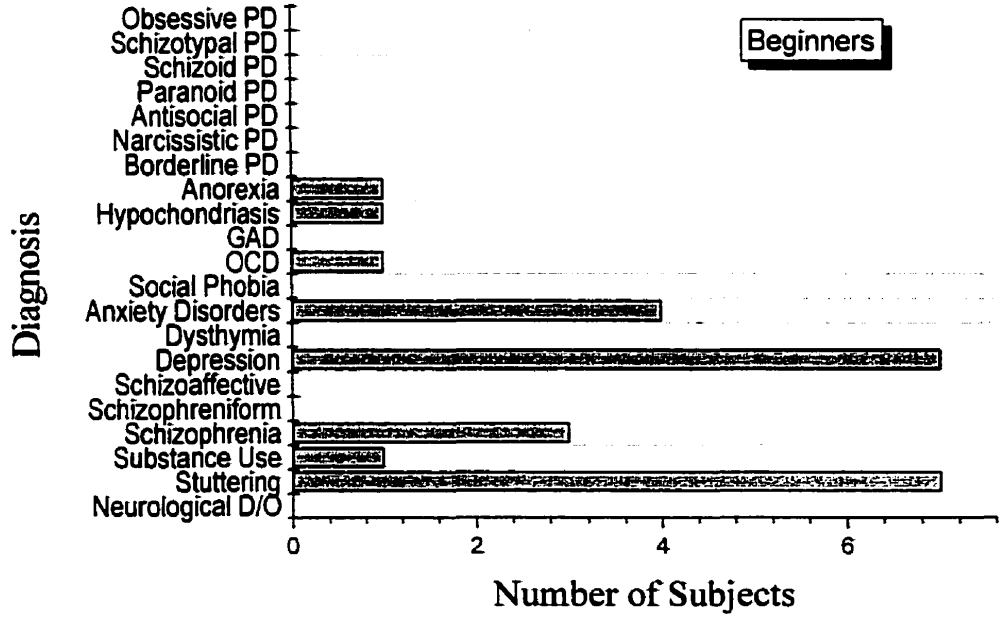


Figure 10 Summary of Diagnoses provided by In-Training Subjects (N=13)

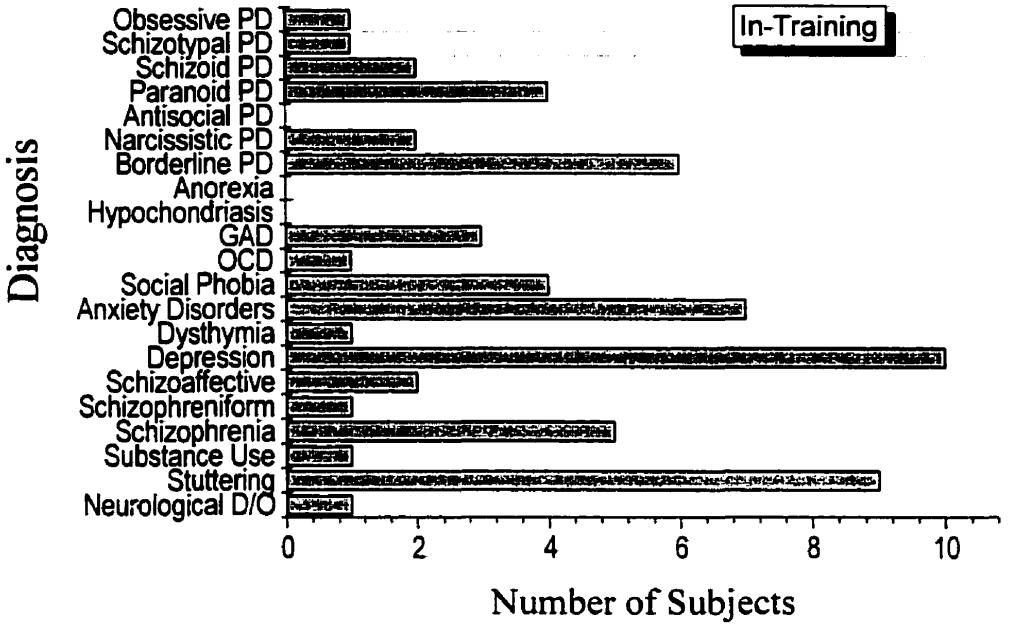
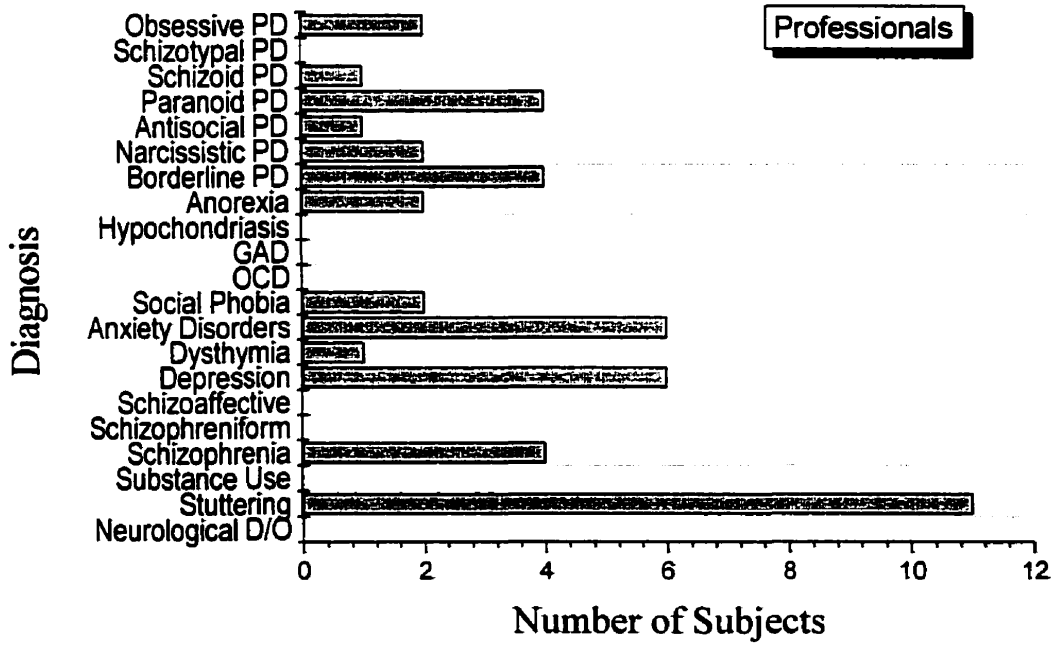


Figure 11 Summary of Diagnoses provided by Professional Subjects (N=14)



In Figure 9, a summary of diagnoses provided by Beginner subjects was presented. Upon examination, one can determine that no Beginners diagnosed Axis II personality disorders any of which would have been considered an accurate diagnosis. The most endorsed Axis I diagnosis was a tie between Depression and Stuttering, and the second most popular diagnosis was Anxiety Disorders. Th B eginners used five other Axis I diagnoses (Substance Use, Schizophrenia, Obsessive-Compulsive Disorder (OCD), Hypochondriasis and Anorexia). Of these five other diagnoses, Substance Use, OCD and Hypochondriasis are not supported by the interview data nor the testing results. This indicates that, collectively, the Beginner group suggested three diagnoses that are not supported by the clinical evidence.

In Figure 10, a summary of diagnoses provided by In-Training subjects was presented. Upon examination, one can determine that many of the In-Training subjects diagnosed Axis II personality disorders, and the most endorsed Axis II personality disorder was Borderline Personality Disorder, with six subjects correctly diagnosing this particular personality disorder. The most endorsed Axis I diagnosis was Depression, the second most popular diagnosis was Stuttering, while the third most endorsed Axis I disorder category was Anxiety Disorders. The In-Training group used nine other Axis I diagnoses (Neurological Disorder, Substance Use, Schizophrenia, Schizophreniform, Schizoaffective, Dysthymia, Social Phobia, Obsessive-Compulsive Disorder (OCD), and Generalized Anxiety Disorder (GAD)). Of these nine other diagnoses, Neurological Disorder, Substance Use, and OCD are not supported by the interview data nor the testing results. This indicates that, collectively, the In-Training group suggested three diagnoses that are not supported by the clinical evidence.

In Figure 11, a summary of diagnoses provided by Professional subjects was presented. Upon examination, one can determine that many Professionals diagnosed Axis II personality disorders, with a tie between Borderline Personality Disorder and Paranoid Personality Disorder for the most endorsed personality disorder. The most endorsed Axis I diagnosis was Stuttering, and the second most popular diagnosis was a tie between Depression and Anxiety Disorders. The Professionals used four

using the DSM-IV Multiaxial Form (Appendix 3 contains a copy of this form). Again, the reasons for using this additional diagnostic extraction measure were that the form prompts for Axis II personality disorders, and that this form is recommended as an aid in conceptualizing diagnoses (APA, 1994).

Table 4.7 (below) presents the pool of accurate diagnoses endorsed by subjects (Axis II personality disorders) across methods of extraction (think aloud (T) versus Multiaxial form (M)). The accuracy or endorsement rates are contained in the cells for each group.

TABLE 4.7 Summary of Accurate Diagnoses provided by subjects Using Think Aloud Method and Multiaxial Form according to level of Experience

	% In-Training N=13		**% Professionals N=14**	
Axis II Personality Disorders offered as the diagnosis	**Think Aloud (T)**	**Axial Form (M)**	**Think Aloud (T)**	**Axial Form (M)**
Borderline	30.7	38.4	21.4	28.5
Antisocial	0.0	0.0	7.1	7.1
Narcissistic	15.3	7.6	7.1	14.2
Paranoid	0.0	30.7	7.1	28.5
Schizoid	7.6	15.3	7.1	7.1
Schizotypal	7.6	7.6	0.0	0.0
Obsessive-Compulsive	0.0	7.6	0.0	14.2

Note 1: Beginner subjects not included in this analysis because none of the subjects correctly detected the presence of an Axis II Personality Disorder

Note 2: There were no significant differences detected between groups on all the above diagnostic accuracy rates

It is interesting to note that in almost every instance, the accuracy rate using the Multiaxial Form (M) is either the same or greater than the accuracy rate using the Think Aloud (T) method. The exception is found in the In-Training group for the diagnosis of Narcissistic Personality Disorder. Borderline Personality Disorder was the most popular diagnosis used by In-Training group as determined by either method of extraction (30.7% for T and 38.4% for M) while there was a tie for the most popular diagnosis amongst the Professional group (28.5% for Borderline PD and 28.5% for Paranoid PD). However, focusing solely on diagnoses provided by Professionals during the think aloud method, Borderline was the most highly endorsed personality disorder by far, with a 21.4% hit rate. Based on these findings, it appears that the use of the Multiaxial Evaluation Report Form (APA, 1994) can function to increase accuracy, at least with regard to Axis II personality disorders. However, use of the Multiaxial form prompts consideration of related diagnoses resulting in accuracy being spread across a larger subset of diagnoses.

Accuracy Cutoffs The second parameter that influences the determination of diagnostic accuracy is the choice of what constitutes an accurate diagnosis. For this study, an accurate diagnosis was conceptualized as consisting of three levels of diagnostic accuracy. The rationale behind this decision was based on the diagnoses obtained from the recommended clinical instruments used to determine the criterion diagnosis (Table 4.6). The three levels, or cutoff points of diagnostic accuracy are Axis II personality disorders, Cluster B personality disorders and Borderline Personality Disorder, while all other diagnoses were considered to be inaccurate for the purposes of this analysis.

Table 4.8 (below) summarizes the changes in diagnostic accuracy according to extraction method (Think Aloud or Multiaxial Form) and as a function of which diagnostic label (Borderline PD, Cluster B PD or Axis II PD) is used as the cutoff. The total cumulative accuracy is the mean accuracy rate of Borderline PD plus additional accuracy gained by using either Cluster B as cutoff or using Axis II PD as cutoff. By expanding the definition of an accurate diagnosis (i.e. making it less

specific), the rate of accuracy increases. Again, use of the Multiaxial Form results in a higher mean accuracy rate for all groups for each of the three cutoffs.

Table 4.8 Summary of Changes in Diagnostic Accuracy as a function of which Diagnostic Label (Borderline PD, Cluster B PD or Axis II PD) is used as cutoff

Diagnostic Label used as Cut off	In-Training Group		Professional Group	
	Think Aloud	Axial Form	Think Aloud	Axial Form
Accuracy (%) using Borderline PD as cutoff	30.7	38.4	21.4	28.5
Total Cumulative Accuracy (%) using Cluster B as cut off (Borderline PD, Histrionic PD, Antisocial PD or Narcissistic PD)	30.7	38.4	35.6	42.7
Total Cumulative Accuracy (%) using Axis II as cut off (Borderline PD, Histrionic PD, Antisocial PD, Narcissistic PD, Paranoid PD, Schizoid PD, Schizotypal PD, Avoidant PD, Dependant PD or Obsessive-Compulsive PD)	38.3	53.7	49.8	78.4

Note 1: Beginner subjects not included in this analysis because none of the subjects correctly detected the presence of an Axis II Personality Disorder

Note 2: There were no significant differences detected between groups on all the above diagnostic accuracy rates

In Table 4.8 (above) the Beginners were eliminated from consideration because none of them provided an Axis II personality disorder diagnosis either during their diagnostic formulations or using the Multiaxial form. From the table, it can be seen that both the In-Training and Professional groups mentioned Axis II diagnoses of varying degrees of diagnostic accuracy. In every instance, use of the Multiaxial Form functioned to increase the collective accuracy of the group. To illustrate, if

definition (small errors in the specificity of their diagnoses); however, they do not appear to make large diagnostic errors by suggesting diagnoses that are inconsistent with the clinical data.

4.5.3 Characteristics of Subjects who provided Accurate Diagnoses

For the purposes of this study, the primary diagnosis of interest was an Axis II Cluster B Personality Disorder called Borderline Personality Disorder. Most subjects provided several diagnoses and many of these were Axis I diagnoses; however, accuracy in this study was determined by inclusion of the diagnosis of interest regardless of other (correct or incorrect) diagnoses provided. Further, accuracy of the primary diagnosis was determined using three cutoffs: Borderline Personality Disorder, Cluster B Personality Disorders and any Axis II Personality Disorders. Therefore, three separate analyses for each method of diagnostic extraction were conducted according to the varying degree of accuracy.

All variables collected in this study were analyzed across each of the accuracy groupings (whether or not subjects diagnosed any Axis II Personality Disorder, any of the Cluster B Personality Disorders and Borderline Personality Disorder). Figure 12 provides a synopsis of significant findings pooled across all analyses.

Figure 12 Characteristics of Participants based on their Diagnostic Accuracy: A Highlight of Significant Findings

Participants who accurately diagnosed:		
Borderline PD[1]	Any Cluster B PD[2]	Any Axis II PD[3]
evidenced the following characteristics		
• Conduct a longer interview • Higher Completeness Index • Greater # verbal exchanges during interview	• Conduct a longer interview • Higher Completeness Index • Greater # verbal exchanges during interview • Higher Test Familiarity Index	• Conduct a longer interview • Higher Completeness Index • Greater # verbal exchanges during interview • Higher Test Familiarity Index
• less likely to ask about presenting complaint • Less likely to ask about work history • less likely to ask about interpersonal relationships • More likely to ask if got diagnosis correct	• More confident about accuracy of their diagnosis • Greater # diagnoses provided in diagnostic network	• Greater # years full time experience • not a Beginner • more likely to specialize in Axis II • Provided a lengthier diagnostic formulation

[1] The diagnosis of interest, belonging to the Cluster B Personality Disorders
[2] Cluster B Personality Disorders includes Borderline as well as Antisocial, Histrionic and Narcissistic Personality Disorders
[3] Any Personality Disorder includes Cluster B Personality Disorders (Borderline, Antisocial, Histrionic, Narcissistic) as well as Cluster A Personality Disorders (Paranoid, Schizoid, Schizotypal) and Cluster C Personality Disorders (Avoidant, Dependent, Obsessive-Compulsive)

Figure 12 (above) illustrates characteristics of participants who correctly diagnosed the case. Using Borderline Personality Disorder as the cut off for an accurate diagnosis, those participants who correctly diagnosed the patient tended

to conduct a longer interview ($F[1,33]=4.829$, $p=0.035$), had a higher Completeness Index ($F[1,31]=8.214$, $p=0.007$), evidenced a greater number verbal exchanges during interview ($F[1,31]=4.582$, $p=0.040$), were less likely to ask about presenting complaint ($F[1,31]=6.293$, $p=0.018$), were less likely to ask about work history ($F[1,31]=4.943$, $p=0.034$), were less likely to ask about interpersonal relationships ($F[1,32]=4.134$, $p=0.050$), and were more likely to ask if they got the diagnosis correct ($\chi^2[1]=3.863$, $p=0.049$). All other variables tested using Borderline Personality Disorder as cut off were not significant.

Using the Cluster B Personality Disorders (Borderline, Antisocial, Narcissistic or Histrionic) as cut off for an accurate diagnosis, those participants who correctly diagnosed the patient conducted a longer interview ($F[1,33]=7.513$, $p=0.010$), had a higher Completeness Index ($F[1,31]=4.972$, $p=0.033$), scored higher on the Test Familiarity Index ($F[1,31]=4.545$, $p=0.041$), were more confident about the accuracy of their diagnosis ($F[1,33]=4.654$, $p=0.038$), evidenced a greater number of verbal exchanges during interview ($F[1,32]=5.790$, $p=0.022$), and provided more diagnoses in their diagnostic networks ($F[1,31]=7.682$, $p=0.009$). All other variables tested using the Cluster B personality disorders as cut off were not significant.

Using any of the Axis II personality disorders as the cut off for an accurate diagnosis (includes Cluster B Personality Disorders (Borderline, Antisocial, Histrionic, Narcissistic) as well as Cluster A Personality Disorders (Paranoid, Schizoid, Schizotypal) and Cluster C Personality Disorders (Avoidant, Dependent, Obsessive-Compulsive)), those participants who correctly diagnosed the patient conducted a longer interview ($F[1,33]=5.862$, $p=0.021$), had a higher Completeness Index ($F[1,32]=4.333$, $p=0.045$), scored higher on the Test Familiarity Index ($F[1,33]=11.865$, $p=0.002$), had a greater number of verbal exchanges during the interview ($F[1,32]=7.120$, $p=0.012$), had more years of full time clinical experience ($F[1,33]=4.238$, $p=0.047$), were not a Beginner ($\chi^2[1]=11.244$, $p=0.004$), were more likely to specialize in personality disorders ($\chi^2[1]=5.042$, $p=0.025$), and provided a lengthier diagnostic formulation ($F[1,33]=8.547$, $p=0.006$). All other

patient, 2) collateral information (information not derived from patient i.e. husband, daughter, coworkers, medical charts), 3) Rorschach, 4) MMPI, 5) MCMI, 6) either TAT or HTP or both, 7) any neuropsychological tests and/or the WAIS, and 8) targeted tests (any test designed to measure a particular construct i.e. Beck for depression, or if they said "a self-esteem test"). The resulting response categories were tabulated. Please refer to Table 4.9 for a summary of tests requested.

Table 4.9 Characterization of tests that were requested by Subjects

Variable of Interest	Level of Experience			Group Differences
	Beginners	In-Training	Professional	
	Mean ± SD (or % yes)	Mean ± SD (or % yes)	Mean ± SD (or % yes)	
Another interview	50	16.7	30.8	$\chi^2[2]=2.527$, p=0.283
Collateral info[1]	1.00±1.31	0.67±0.98	0.31±0.63	F[2,30]=1.336, p=0.278
Rorschach	0	25	23.1	$\chi^2[2]=2.362$, p=0.307
MMPI	0	75	53.8	$\chi^2[2]=11.057$, p=0.004 [a,b]
MCMI	0	25	15.4	$\chi^2[2]=2.334$, p=0.311
TAT/HTP	0	33.3	7.7	$\chi^2[2]=5.077$, p=0.079
Neuropsych /WAIS	12.5	25	7.7	$\chi^2[2]=1.512$, p=0.470
Targeted tests[2]	1.75±1.16	0.92±1.38	0.62±0.87	F[2,30]=2.471, p=0.102

[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

[1] Collateral info is information sought from a source other than the patient (meet with husband, daughter, coworkers, medical records, her GP...)
[2] Targeted tests are measures that target specific areas of functioning (Beck/BDI, anxiety measures, life events questionnaire, self-esteem measure...)

Table 4.9 illustrates that there is virtually no difference between levels of clinical experience and the types of information and tests subjects request. The only significant difference was found for the variable of requesting the MMPI ($\chi^2[2]=11.057$, $p=0.004$). Post-hoc tests revealed that the In-Training and the Professional groups were more likely than the Beginner group to request the MMPI. There were no significant difference between the In-Training and the Professional group to request the MMPI.

In addition to the specific breakdown of tests that were requested (Table 4.9), additional information on the use of the testing material was analyzed. Whether subjects requested any tests and the number of tests requested was tabulated. Furthermore, an unexpected phenomenon occurred during the assessment process; namely, that the majority of subjects incorporated their diagnostic impressions when asked for additional tests they would consider in the evaluation of the simulated patient. Therefore, whether or not subjects incorporated diagnostic formulations during this component of the assessment task was determined. After subjects requested tests and additional sources of information, each subject was provided with the same test results, the WAIS, a measure of intelligence, the Rorschach Inkblot Test, a projective test, and the MMPI, a personality inventory. It is interesting to note that, of those subjects who requested tests, 0% of the Beginners, 82% of the In-Training group and 73% of the Professionals requested at least one of either the WAIS, the MMPI and/or the Rorschach. This indicates that in the majority of cases of those with knowledge of psychological testing, at least one of the tests that they requested was given to them to interpret.

Additional analyses included the examination of whether subjects referred to any of the psychometric tests during their diagnostic formulations. A comparison of diagnoses provided during the "Request Tests" section with diagnoses provided during the "Diagnostic Formulation" section permitted an examination of the effect of introduction of psychometric test data on accuracy of diagnostic impressions. This analysis was limited to looking at whether subjects made an incorrect diagnosis, an Axis II personality disorder diagnosis or a Borderline Personality Disorder diagnosis. An increase in accuracy was rated if a subject went from inaccurate

diagnosis to Axis II or Borderline diagnosis, or if subjects' diagnoses went from Axis II to Borderline.

Finally, whether subjects incorporated treatment recommendations into their diagnostic formulations was examined. This was motivated by the observation that subjects tend to structure the assessment task according to how they conduct their clinical activities, rather than how the researcher decided to structure the assessment task. In other words, clinical activities such as seeking additional information ("Request Tests"), diagnostic formulations, and therapeutic recommendations tend to cluster together in the clinicians thinking, rather than being clearly articulated during specific components of the assessment task. Thus, when analyzing data according to task, one might miss information contained in other sections of the task. For example, the diagnostic information contained in the "Request Tests" section may be lost if one limits the examination of diagnostic considerations obtained solely during the diagnostic formulation component of the assessment task. Table 4.10 summarizes the results of the above analyses.

as to whether they asked for tests or additional information (χ^2[2]=1.431, p=0.489), nor was there a significant difference between groups on the number of tests or measures they requested (F[2,30]=1.17, p=0.325). The tendency to verbalize diagnostic impressions during the "Request Tests" section was greater for the In-Training and the Professional groups (χ^2[2]=5.50, p=0.064), with one subject from the Beginner group compared to the majority of subjects from the groups with clinical training and experience. Post-hoc analyses indicated a significant difference between the Beginners and the other groups on this measure. There was no significant difference between groups in terms of referring to the psychometric tests (WAIS, MMPI or Rorschach) during their diagnostic formulation. There did not appear to be an effect of psychometric test data on increased diagnostic accuracy (χ^2[2]=3.402, p=0.183). Finally, post-hoc analyses revealed that significantly more Professionals (61.5%) than any other group (12.5% of Beginners or 25.0% of the In-Training group) incorporated treatment considerations during the "Diagnostic Formulation" section (χ^2[2]=6.199, p=0.045). This might indicate a tighter coupling between diagnosis and treatment considerations for professional psychologists. This finding parallels the demands of clinical activities of private practitioners in clinical psychology. Private practitioners' primary clinical function is to treat their clients/patients. Thus, when conducting an assessment, it is not surprising that they are weaving treatment considerations into their understanding of the patient, because whatever the diagnosis may be, it serves to inform psychotherapeutic behaviour and expectations.

4.7 Subjective Variables

Subjective variables were sought to capture subjects' reactions about their performance and the study. Confidence about their diagnoses was one of these subjective variables, as well as how realistic, or ecologically valid they found the diagnostic portion of the study. Both confidence and ecological validity were assessed using a Likert-type rating scale. For the confidence question, subjects were asked "How confident are you in your judgment for this case?" with one being "very unconfident" and seven being "very confident". For the ecological validity of the

study, subjects were asked to "Rate the ecological validity of the diagnostic portion of the study" with one being "no resemblance to real assessments" and seven being "as realistic as can be". Finally, whether subjects voluntarily asked if they correctly diagnosed the case was recorded. Asking may imply an effort to seek feedback, curiosity, or perhaps some uncertainty as to their performance. Please refer to Table 4.11 for a summary of findings and group means.

Table 4.11 Summary of Results for Subjective Variables

Variable of Interest	Level of Experience			Significance
	Beginners Mean ± SD (or %)	In-Training Mean ± SD (or %)	Professional Mean ± SD (or %)	
Confidence Rating (1 to 7)	4.13 ± 1.13	3.77 ± 1.17	4.71 ± 0.83	F[2,32]=2.882, p=0.071
Validity Rating (1 to 7)	5.25 ± 1.16	5.23 ± 1.17	5.43 ± 1.09	F[2,32]=0.119, p=0.888
Did they Ask if Diagnosis Correct? (% Yes)	100 % (8 out of 8)	69.2 % (9 out of 13)	50 % (7 out of 14)	χ^2[2]=5.910, p=0.052 [b]

[a] Beginner and In-Training groups significantly different from each other
[b] Beginner and Professional groups significantly different from each other
[c] In-Training and Professional groups significantly different from each other

In Table 4.11 three subjective variables were presented, and a one-way ANOVA was used to analyze differences in confidence and validity ratings. The Analysis of Variance failed detected a trend for the variable of confidence rating (F[2,32]=2.882, p=0.071). It appears that the In-Training group reported having the least amount of

analyses were used to look at differences between groups using all variables previously examined in the study. Analyses revealed a handful of variables that differentiate between the two groups, with the group with extensive experience (Professional 25 years) tending to be: older ($F[1,12]=11.254$, $p=0.006$), have a greater number of years full time experience ($F[1,12]=58.685$, $p=0.000$), asked proportionately fewer questions about current living situation ($F[1,12]=12.326$, $p=0.004$), asked proportionately more questions about interpersonal relationships ($F[1,12]=5.390$, $p=0.039$), evidenced a lower Completeness Index of the interview ($F[1,12]=9.000$, $p=0.011$), had fewer pieces in the diagnostic network ($F[1,12]=10.323$, $p=0.008$), and had fewer chunks in the diagnostic network ($F[1,12]=6.471$, $p=0.027$). A significant difference was not detected on all other variables studies, including accuracy measures. This indicates that extensive experience (about 25 years) does not appear to increase the rate of diagnostic accuracy as measured in this study. Please refer to Table 4.12 (next page) for a summary of significant findings.

Table 4.12 Summary of Significant Findings between the Professional group with 10 years experience and the Professional with 25 years clinical experience

Variable of Interest	Degree of Clinical Experience		Statistic
	Professionals with 10 years clinical experience (n=7)	Professionals with 25 years clinical experience (n=7)	
	x ± SD	x ± SD	
Age	44.43 ± 6.75	56.43 ± 6.63	F[1,12]=11.254, p=0.006
Years full time experience	11.86 ± 2.73	27.86 ± 4.80	F[1,12]=58.685, p=0.000
Proportion of questions about current living situation	0.010 ± 0.007	0.000 ± 0.000	F[1,12]=12.326, p=0.004
Proportion of questions about interpersonal relationships	0.186 ± 0.106	0.486 ± 0.354	F[1,12]=5.390, p=0.039
Completeness Index	11.71 ± 1.89	8.28 ± 2.36	F[1,12]=9.000, p=0.011
Pieces in diagnostic network	4.33 ± 2.34	1.43 ± 0.53	F[1,12]=10.323, p=0.008
Chunks in diagnostic network	2.33 ± 1.63	0.57 ± 0.79	F[1,12]=6.471, p=0.027

Table 4.12 (above) summarizes the variables that significantly differentiate the performance of subjects according to degree of extensive clinical experience. The more experienced group of psychologists covered fewer interview topics during the clinical interview (8.28 out of 18) compared to (11.71 out of 18) for the psychologists with ten years experience. This indicates a less comprehensive array of interview topics covered by individuals with extensive clinical experience. However, from the topics that they do ask, psychologists with 25 years experience tended to ask significantly more about interpersonal relationships, indicating that they feel this topic to be especially fruitful in terms of clinical information gained. One explanation as to why fewer topics are covered in the clinical interview might be that psychologists with extensive experience do not benefit from conducting comprehensive clinical interviews because they are already focusing on topics that they find most informative to their diagnostic considerations. However, this study determined that a higher Completeness Index is a characteristic of subjects who provided an accurate diagnosis. It may be that those subjects who do not have extensive experience benefit most from conducting a comprehensive interview in order to maximize diagnostic accuracy, but that psychologists with extensive experience can diagnose just as accurately, without having to cover as many topics. However, the effect of covering all recommended topics in psychologists with extensive experience might function to enable them to increase their diagnostic accuracy beyond that of less experienced subjects, a hypothesis that warrants further investigation.

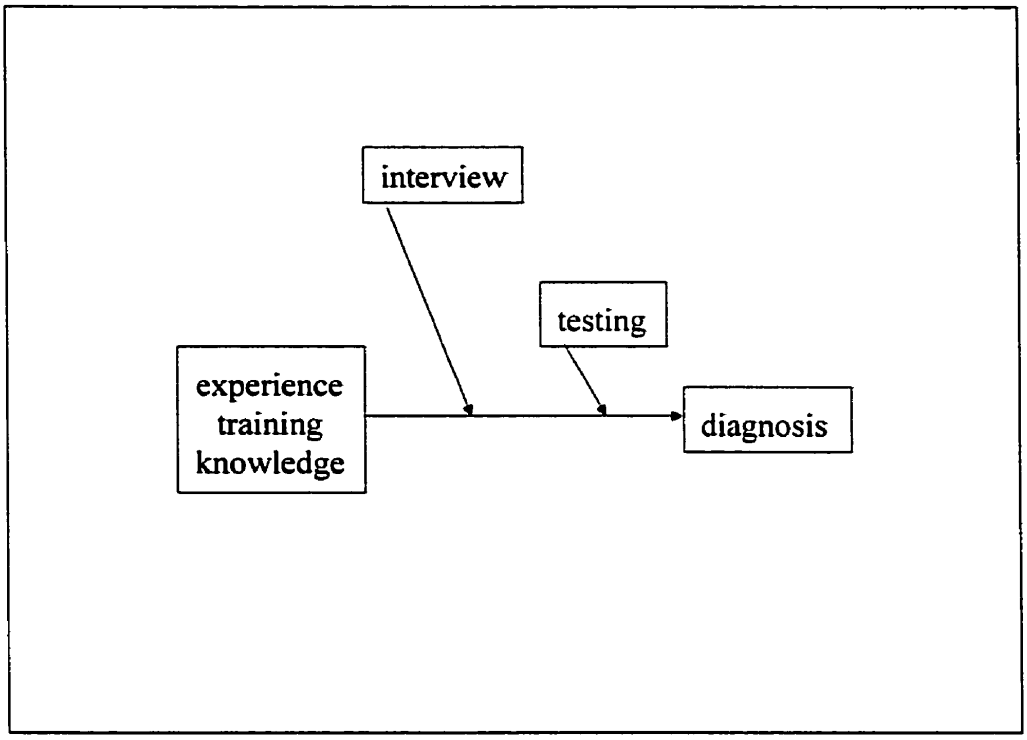
The final significant difference that deserves mention is that psychologists with extensive experience evidence fewer pieces (1.43) and chunks (0.57) in their diagnostic networks compared to psychologists with about ten years clinical experience (4.33 pieces and 2.33 chunks). This finding is consistent with the Professional group demonstrating a tendency toward formulating cohesive, interconnected diagnostic networks. It may be that extensive experience enables a clinician to better conceptualize newly acquired information about a patient into their diagnostic considerations as a fully integrated whole, with clinical concepts and important pieces of information conceptually linked.

Although this thesis sought to investigate diagnostic accuracy using diagnostic entities found in the current classification system for mental disorders, the DSM-IV, and comparing the diagnostic label generated by participants to a criterion diagnosis, it may be that the more fundamental understanding any patient, whether correctly labeled or not, might be captured in the networks. These networks tend to be more internally cohesive in those with more experience. Thus, while we can say that there is no difference in diagnostic accuracy as a function of extensive experience, one might suggest that there are increases in clinical case comprehension and understanding with increasing clinical experience.

4.9 Proposed Model to Characterize the Assessment Process

In order to organize some of the major findings generated from the thesis, a model of the assessment process, adapted from the medical decision-making domain, was developed. The proposed model is knowledge-based in that it structures the assessment process in terms of the organization and availability of knowledge available to the clinician (Keravnou & Johnson, 1989; Groen & Patel, 1985; Bordage, Grant & Mardsen, 1990) as well as what information sources drives the diagnostic formulation. The knowledge-based model can be contrasted with the hypothetico-deductive model whose main components are data acquisition, hypothesis generation, data interpretation, and hypothesis evaluation (Elstein, Shulman & Sprafka (1978); Barrows, Feightner, Neufeld & Norman (1978)). The hypothetico-deductive model proved inadequate in identifying expert and non-expert diagnosticians (Bordage, Grant & Mardsen, 1990), and thus was not used. First, an example of the model typically used in clinical decision-making research in psychology is presented in Figure 13 .

Figure 13 Model of a Traditional Assessment Task found in Clinical Psychology



In this typical assessment model, the clinical knowledge, training and experience is varied in order to determine its influence on diagnostic accuracy. Usually, case material from an interview, and/or testing is pre-selected and presented to subjects. Subjects are then asked to use this data in their diagnostic formulation or judgment tasks. The design implies a control over data, where subjects are provided with the same data for combination. Thus, differences in diagnostic accuracy are deemed a function of clinical experience, since the data was held constant. Studies that follow this model typically determine that increasing clinical experience does not increase diagnostic accuracy (See Dawes, 1994, for a review). What this model fails to consider is the influence of the clinical knowledge, training and experience on the structuring and acquisition of interview and/or testing data.

Figure 14 (below) presents a modified model that incorporates the influence of previous knowledge, training and experience in the assessment process from beginning to end. The proposed model serves to more accurately capture and represent the assessment process of subjects in this study.

testing materials were sources of clinical information. Second, subjects' training, experience and knowledge shaped the amount and nature of the information extracted. Third, the clinical information was superimposed on a preexisting clinical knowledge structure, which, in turn, was affected by training and experience.

Figure 15 (below) presents a tabulated summary of the significant findings in this study as a function of clinical experience during the assessment process, where findings are organized into components meant to represent aspects of the proposed model. For instance, the nodes of "Knowledge", "Interview", "Testing" and "Diagnosis" in the proposed model are captured by the corresponding headings in Figure 15. Further, nonsignificant findings that are typical of all subjects are presented to illustrate how and in what manner subjects' clinical behaviour is both similar and different during the assessment process.

Figure 15 The Assessment Process as a Function of Clinical Experience

Level of Experience	Clinical Knowledge	Interview	Testing	Diagnosis
ALL levels (variables where subjects did not differ in regards to their assessment behaviour)		• same # encouraging statements • many of topics were covered by all groups • same time spent interviewing	• request at least one test • refer to tests during dx to same degree	• same # diagnoses provided • accuracy increased with multiaxial form • no difference in confidence, validity ratings
Beginners	• Lowest test familiarity index • Does not specialize in personality disorders • No clinical experience • no training	• Fewest # verbal exchanges • greatest # close ended questions • Never asked about suicide attempts or mental status	• 0% request MMPI	• Most likely to ask if got dx correct • Did not correctly diagnose case • Least likely to provide dx during request tests
In-Training	• some clinical experience (2 years) • various levels of training	• greatest # open ended questions • Most likely to ask about leisure activities and disorder specific questions • Highest Completeness Index	• 75% request MMPI	• accuracy varied from 31% to 54% depending on cutoff • provide some inconsistent dx • 58% provide dx during request tests
Professional	• Works greatest # of clinical hours per week • most clinical experience (20 years) • fully trained	• similar levels of open & close-ended questions asked during interview	• 54% request MMPI	• took longest to provided diagnostic formulation • accuracy varied from 21% to 78% depending on cutoff • all diagnoses consistent with data • least likely to ask about diagnosis

CHAPTER V

GENERAL DISCUSSION & CONCLUSIONS

The goals of this thesis were to 1) examine the assessment process in psychology as a function of clinical experience, 2) to examine characteristics of participants who provided an accurate diagnosis, 3) to develop models of the clinical assessment process in psychology as a function of experience, and 4) to explore the effect of extensive clinical experience on the assessment process. These goals have been achieved in the preceding Results and Discussion chapter, but highlights of significant findings are presented below. Limitations to the present investigation follow the summary, as well as a statement on the contribution to knowledge.

5.1 Synopsis of top level findings

1 - Clinical Training and Familiarity with Psychological Testing Material The general clinical preparation between the groups in this study was examined. The Professional group evidenced the highest amount of experience and clinical activities, but did not differ from the In-Training group on their familiarity with psychological testing.

One interpretation of the lack of difference between the In-Training group and the Professional group on their familiarity with popular psychological tests (see Appendix 2 for the tests comprising the index) is that the majority of training in psychological testing is accomplished during graduate school. Given that increasing clinical experience does not equate with a greater familiarity of psychological tests, exposure to and training in psychometric test administration and interpretation must be provided at the graduate level.

Alternatively, it could be suggested that most of the tests and measures deemed important to master consist of instruments developed decades ago. Thus, clinical

experience might not increase test familiarity because although versions are being revised, new tests are not being developed. This was exemplified in a recent survey of clinical training directors who rated clinical tests and measures they felt were important that psychology interns be familiar with. Among the top instruments were the Rorschach Inkblot test, the Thematic Apperception Test (TAT), the Minnesota Multiphasic Inventory (MMPI) and the Weschler Adult Intelligence Scale (WAIS) (Clemence, & Handler, 2001). The training directors prefer that interns are familiar with the latest versions of these instruments; however, the introduction of many of these instruments dates far back. For instance, the Rorschach was developed in 1921, and the WAIS was first developed in 1955, although there have been two subsequent revisions.

Based on this particular finding, it is recommended that adequate exposure to and training in the use of psychometric test materials during graduate school be maximized for this appears to be a critical period during clinicians' training. If clinicians are not taught the administration and interpretation of particular tests, it does not appear that increasing amounts of clinical experience would remedy the lack of knowledge.

2 - Time and Longer Diagnostic Formulations The length of time subjects spent completing the various components of the assessment task was examined. It is believed that the longer it takes for subjects to complete components of the assessment task, the greater importance and effort is devoted to that component. The Professional group took almost double as long to complete their diagnostic formulations, indicating an emphasis on this portion of the assessment task. The additional time may be spent conceptualizing, producing, elaborating and refining diagnostic hypotheses.

3 - Differing Interview Strategies Developmentally, there appear to be stages or predictors that characterize subjects' performance during the interview. Those with the least clinical experience rely on close-ended questions during the clinical interview. It might be that graduate departments are aware of this initial reliance on

specific questioning, which then results in increased diagnostic accuracy. By following these kinds of recommendations from the literature, this group was just as likely to provide an accurate diagnosis as their comparison group with years of clinical experience.

On the other hand, despite lower Completeness Index scores, the professional psychologists were nevertheless able to obtain statistically equivalent accuracy rates as those with higher Completeness Index scores. This suggests that the Professionals are more efficient in targeting relevant information needed for the diagnostic conceptualization. It might be that they delve deeper into the pertinent topics, while leaving enough time to spare. With the extra time, they may provide information to the patient potentially resulting in a therapeutic effect. A future study might compare a group of professionals conducting an assessment using their usual strategies with a group of professionals using comprehensive interviewing with diagnostic criteria-specific questions. It remains to be seen whether experienced professionals could increase their diagnostic accuracy with a targeted, comprehensive interviewing strategy.

5 - Diagnoses Consistent with Clinical Data Although both the In-Training and Professional groups provide accurate diagnoses amongst their pool of tentative diagnoses, the In-Training group also included diagnoses that are inconsistent with the clinical data. This implies that those with less clinical experience are capable of diagnostic accuracy at the same rate as those with much more experience; however, they are also more likely to propose diagnoses that are totally inaccurate and that are not supported by the clinical evidence. The Professionals, however, also provide additional diagnoses, but the nature of these diagnoses is reasonable given the clinical data. These findings suggest that experience plays a role in narrowing the focus of diagnostic accuracy to a more relevant level. Those with significantly less experience cast wider diagnostic nets and capture the diagnosis of interest, but included amongst those diagnoses are categorically wrong diagnoses. It appears that experience affords the clinician the ability to avoid large diagnostic errors.

6 - Diagnostic Accuracy and the Use of the Multiaxial Evaluation Form The use of the Multiaxial Form is highly recommended, especially in the detection of Axis II presence, as it increased the accuracy rate of clinicians (for each cutoff diagnosis) when compared to using just the Think Aloud method.

7 – Characteristic of those with Accurate Diagnoses When comparing the characteristics of those who provided accurate diagnoses to those who did not – regardless of which cutoff diagnosis was used, and regardless of level of experience – we find that those who were accurate tended to conduct longer interviews and demonstrated a greater number of verbal exchanges between themselves and the patient. This indicates that the increased time spent interviewing the patient is due to a greater activity of questioning for a longer period of time rather than simply conducting a longer interview. Further, conducting a thorough and comprehensive interview in terms of covering recommended topics (Morrison, 1995) is related to accuracy, at least with Axis II personality disorders. Familiarity with psychometric testing is related to increased diagnostic accuracy. This may occur due to increments in relevant clinical information that testing results may contribute to the assessment process. Finally, seeking feedback as to performance can function to refine diagnostic specificity by allowing clinicians to update their knowledge and alter their assessment strategies in order to increase diagnostic accuracy and specificity.

8 - The Enmeshment of the Clinical Assessment Process Subjects, especially those with at least some clinical experience, tended to integrate components of the assessment task together, rather than limiting themselves to responding within the constraints of the task at hand. For example, many of the subjects from the In-Training and Professional groups incorporated diagnoses in their "Request Tests" section. In a similar vein, Professionals incorporated therapeutic considerations into their diagnostic formulations. This finding might imply that subjects with clinical training and at least some clinical experience combine the major components of the assessment process in a parallel manner, rather than the sequential method employed

in this study. The preceding results support the notion that certain components of the assessment process, namely consideration of testing, diagnostic formulation and treatment prescription are integrally tied in actual practice, and that research into the assessment process should take this into consideration for design of studies.

9 - Extensive Experience and no change in Diagnostic Accuracy Those with extensive experience tend to have scored lower on the Completeness Index, indicating that they cover fewer topics during their clinical interview with the patient, but they still attain the same level of accuracy as those who conduct comprehensive interviews. An explanation might be that psychologists with extensive experience do not benefit from conducting comprehensive clinical interviews because they cover topics that are most important to their diagnostic considerations.

Those with extensive experience also show a tendency to provide fewer pieces or chunks in their diagnostic networks, indicating an increased cohesiveness of case comprehension. This might have implications for treatment prescriptions or the quality of therapy with more experienced psychologists.

The final difference between psychologists with ten years versus 25 years clinical experience is that those with decades of experience ask more about interpersonal relationships during the interview. This indicates a greater emphasis placed on this kind of information, and that with increasing experience, psychologists will focus on relational functioning of the patient with others in the patient's life. It may be that this topic can best characterize the nature of the person's difficulties and provide the clinician with the most useful and informative information about the patient.

10 – A Descriptive Model of Clinical Assessment Behaviour A model of the assessment process in clinical psychology was proposed, partly based on the study's findings and partly based on the inaccuracies of the traditional model. It suggests that clinical knowledge, experience and training function to influence the accumulation of clinical data about a particular patient. Studies investigating the effect of experience merely investigate one role that experience plays in the

assessment process, namely the combination of clinical data. The proposed model suggests that experience, training and clinical knowledge impact the amount, nature and kind of clinical data gleaned from an encounter with a patient. It is recommended that future studies incorporate this model as a way to conceptualize the assessment process.

5.2 Limitations and Future Research

In the design of any study, there are decisions made which influence data collection and subsequent analysis, and these choices are informed by the goals of the study. The choices made in this study, in attempting to attain objectives within the constraints encountered, put a number of limitations into play. The first limitation is in regard to the relatively small sample size. A relatively small sample of thirty-five subjects allowed detailed analysis of individual protocols. As the sample size in this study was informed by power and effect size calculations, it was determined that differences detected between groups would represent a large effect size. Thus, increasing the number of subjects in each experience grouping would enable the detection of more subtle influences on the assessment process. For example, some of the trends reported in the thesis might prove to be significant with larger sample sizes.

Another potential inadequacy of the study was that, although great efforts were taken to realistically portray a patient, a simulated patient is not the same as an actual patient. It could be that having an actual patient to interview provides clinicians with information not captured by simulations. Thus, future studies might extend this investigation to include actual patients to study what information experienced clinicians seek when interviewing real patients/clients, and how the lack of such information, as in this study, may affect the clinician's effectiveness. This way one can study the extent to which theory is applicable in the naturalistic situation. It should be stated however, that the use of a trained simulated patient does have the advantage of keeping the delivery of information and interview behaviour as constant as possible across all subjects. This ensures that the same question from

different clinicians would be answered in the same manner. Thus, future studies incorporating actual patients must train the patients to behave in a consistent manner across all subjects.

The researcher was not blind to subjects level of expertise because she portrayed the patient for every subject. This might be remedied in the future by having a second person blind to subjects' level of experience portraying the patient. However, one might still consider a younger subject to belong to a less experienced group, while a more senior individual would in all likelihood be considered to belong to the clinically experienced group. Another related limitation involves the reliability or potential increase in fidelity of portraying the patient over time. In other words, perhaps the first few portrayals might not have been as natural or as well rehearsed as the last few portrayals. This might imply that the subjects might have received different information due to differences in practice effects of the researcher portraying the patient. Although this might be a concern, this study ran subjects in a random manner so that there were subjects from all levels of experience being run at varying degrees of practice in portraying the patient.

The groupings of subjects were based on their clinical experience, and differences in performance across the variables of interest were attributable to differences in levels of clinical experience. However, other variables might account for differences in performance such as the age of the subject or the differences in training methods used for different cohorts of psychologists. To tease apart the effect of age on performance, future studies might employ older subjects who have recently commenced their training in clinical psychology and match ages with psychologists with extensive clinical experience. To elucidate the effects of training regimens and amount of clinical experience, future studies might compare the same cohorts of psychologists, but divide them according to exposure to clinical practice. For example, compare psychologists trained using similar methods at similar times that practice full time versus those who practice part time or less.

A final limitation of the present study results from the attempt to maintain an ecologically valid sampling of assessment behaviour. This thesis sought to investigate the overall assessment process and therefore methods of data collection

of this investigation attempted to minimize interference with this process. Additional probes, or explanation tasks could be employed in future studies to more fully examine specific aspects of this process by asking subjects about their assessment behaviour, and justifications as to why certain strategies were employed.

Future studies in the area could expand the methodology used in this study, and incorporate additional methods of analysis in order to explore the assessment process in more depth and detail. For example, videotaping initial interviews with actual patients then probing the psychologist immediately afterwards for a formulation. Later, the researcher and psychologist could view the videotape and the psychologist could explain why s/he asked various questions and what they remember thinking at the time, although biases in the recall of this kind of information would have to somehow be controlled for.

As researchers learn more about the science of psychopathology and the behaviour of clinicians, methods will adapt to capture knowledge of the assessment process. The ultimate goal is to understand behaviour so that recommendations for improvements in training and assessment behaviour can be informed by data on how psychologists *actually* assess patients.

5.3 Summary Statement – Contribution to Knowledge

> Of all patients with mental disorders, about 50% are treated by primary-care physicians, 25% by mental health professionals and an additional quarter go untreated
>
> Maxmen & Ward (1995)

Our understanding of mental disorders has increased dramatically, yet there remains decades of research before this understanding can attain the level that medicine has achieved regarding physical disease. However, our population requires effective treatment today, rather than sometime in the distant future.

By investigating what many believe to be a critical step in the effective treatment of patients with mental disorders - the assessment process - this thesis explored the entire endeavor employing an ecologically valid design and using proven methods of

analysis in order to accurately capture what psychologists do during an assessment. Adapting methods of analysis frequently utilized in cognitive science, but not widely embraced in investigating clinical decision-making in psychology, permitted a novel method of investigation of the assessment process in clinical psychology. In this way, the design of this project attempted to remedy some of the limitations found in previous studies on the examination of clinical decision-making in psychology.

This thesis contributed to knowledge in psychology by generating a number of findings. As also found in the literature, this study found there to be similar accuracy rates for graduate students in clinical psychology and experienced psychologists. However, experienced psychologists proposed diagnoses that were consistent with clinical data, whereas those with less clinical experience were more likely to make errors by proposing diagnoses that are not consistent with clinical data. Another finding indicates that experienced psychologists evidence similar rates of open and close-ended questions during the interview, while those with no training demonstrate a preponderance of close-ended questions and those with some training evidence more open-ended questions. This indicates a developmental trajectory of interviewing styles that is highly influenced by level of clinical experience. Finally, graduate students in clinical psychology conduct comprehensive interviews while experienced psychologists tend to inquire about fewer topics, but retain the same degree of accuracy. This might imply different strategies based on level of experience and clinical knowledge. These findings, and the others discussed earlier, added to and enriched our knowledge about how psychologists conduct assessments.

The final contribution this thesis makes to the advancement of science is the proposal of a model for investigating assessment behaviour. The proposed model more accurately captures clinical behaviours during the assessment process. The proposed model could be employed and adapted in future research as a framework to investigate complex decisions in other healthcare fields such as nursing, medicine, social work or any other domain that investigates the decision-making process in complex environments.

Beach, L.R. & Lipshitz, R. (1993). Why Classical Decision Theory is an Inappropriate Standard for Evaluating and Aiding Most Human Decision-making. In G.A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), *Decision-making in Action: Models and Methods,* (p. 21-35). Norwood, New Jersey: Ablex Publishing Corporation.

Blais, M.A., Hilsenroth, M.J. & Castlebury, F.D. (1997). Content validity of the DSM-IV borderline and narcissistic personality disorder criteria sets. *Comprehensive Psychiatry. 38(1),*31-37.

Blois, M. (1988). Medicine and the Nature of Virtual Reasoning. *The New England Journal of Medicine, 318(13),* 847-851.

Bockenholt, U. & Weber, E. (1992). Use of Formal Methods in Medical Decision-making: A Survey and Analysis. *Medical Decision-making, 12(4),* 298-306.

Bordage, G., Grant, J. & Marsden, P. (1990). Quantitative Assessment of Diagnostic Ability. *Medical Education, 24,* 413-425.

Borgade, G., Lemieux, M. (1991). Semantic Structures and Diagnostic Thinking of Experts and Novices. *Academic Medicine, 66(9),* S70-S72

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45,* 223-241.

Camerer, C.F. & Johnson, E.J. (1991). The Process-Performance Paradox in Expert Judgment: How can experts know so much and predict so badly? In K.A. Ericsson & J. Smith (Eds.), *Towards a General Theory of Expertise: Prospects and Limits,* (pp.195-217). New York, NY: Cambridge University Press.

Caplan, P.J. (1995). *They Say You're Crazy: How the World's Most Powerful Psychiatrists Decide Who's Normal.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Chase W.G. & Simon H.A. (1973). Perception in chess. *Cognitive Psychology, 1,* 55-81.

Cimino, J.J. (1999). Development of Expertise in Medical Practice. In R.J. Sternberg & J.A. Horvath (Eds.), *Tacit Knowledge in Professional Practice,* (p. 101-120). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Clark, L.A., Watson, D., & Reynolds, S. (1995). Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annual Review of Psychology, 46,* 121-153.

Clemence, A.J. & Handler, L (2001). Psychological assessment on internship: A survey of training directors and their expectations for students. *Journal of Personality Assessment, 76(1),*18-47.

Coles, E. M. (1996). A comment on DSM-IV'S diagnostic taxa. *Psychological Reports. 78(2),*688-690.

Davis, G.C. & Akiskal, H.S. (1989). Descriptive, biological, and theoretical aspects of borderline personality disorder. *Hospital & Community Psychiatry.37(7),* 685-692.

Dawes, R.M. (1988). *Rational Choice in an Uncertain World.* Fort Worth: Harcourt Brace College Publishers.

Dawes, R.M. (1994). *House of Cards: Psychology and Psychotherapy Built on Myth.* New York: The Free Press, A Division of Simon & Schuster, Inc.

Dawes, R.M., Faust, D. & Meehl, P.E. (1989). Clinical Versus Actuarial Judgment. *Science, 243,* 1668-1674.

DeKleer, J. & Brown, J.S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A.L. Stevens (Eds.), *Mental Models.* Hillsdale, NJ:Erlbaum.

Dienst, E.R. & Armstrong, P.M. (1988). Evaluation of Students' Clinical Competence. *Professional Psychology: Research and Practice, 19(3),* 339-341.

Dino, G.A., Shanteau, J., Binkley, M., & Spenser, A. (1984). The detrimental effects of environment stress on creativity. *Technical Report No, 84-2.* Kansas State University, Department of Psychology.

Doane, S.M., Pellegrino, J.W., & Klatzky, R.L. (1990). Expertise in a computer operating system: Conceptualization and performance. *Human-Computer Interaction, 5,* 267-304.

Dunne, D. & Chute, C.G. (1999). Detailed content and terminological properties of DSM-IV. *Proceedings, AMIA Annual Symposium,* 57-61.

Eddy, D.M. (1988). Variations in Physician Practice: The Role of Uncertainty. In J. Dowie & A. Elstein (Eds.), *Professional Judgment A Reader in Clinical Decision-making,* (p. 45-59). Cambridge, MA: Cambridge University Press.

Eddy, D.M. & Clanton, C.H. (1982). The Art of Diagnosis: Solving the Clinicopathological Exercise. *The New England Journal Of Medicine, 306,* 1263-1268.

Ericsson, K.A. & Smith, J. (1991). Prospects and Limits of the Empirical Study of Expertise: an Introduction. In Ericsson, K.A. & Smith, J. (Eds.), *Towards a General Theory of Expertise,* (p.1-38). Cambridge University press.

Eyde, L.D. Childs, R.A. (2000). Qualifications of Assessment Faculty in Clinical Psychology Doctoral Programs: An Analysis of American Psychological Association Membership Data. *Professional Psychology: Research and Practice, 31(2),* 165-169.

Faust, D. & Nurcombe, B. (1989). Improving the Accuracy of Clinical Judgment. *Psychiatry, 52,* 196-209.

Follette, W.C. & Houts, A.C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM. *Journal of Consulting & Clinical Psychology. 64(6),* 1120-1132.

Frederiksen, C.H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7,* 371-458.

Gambrill, E. (1990). *Critical Thinking in Clinical Practice.* San Francisco: Jossey-Bass Publishers.

Garb, H.N. (1998*). Studying the clinician: Judgment research and psychological assessment.* Washington, DC, USA: American Psychological Association.

Gartner, J., Hurt, S.W. & Gartner, A. (1989). Psychological test signs of borderline personality disorder: A review of the empirical literature. *Journal of Personality Assessment. 53(3),* 423-441.

Gill, M., Newman, R. & Redlich, F.C. (1954). *The Initial Interview in Psychiatric Practice.* New York: Hallmark-Hubner.

Glaser, R. & Chi, M.T.H. (1988). Overview. In Chi, M.T.H., Glaser. R. & Farr, M.J. (Eds.), *The Nature of Expertise,* (p. xv-xxviii). Lawrence Erlbaum Associates.

Glidewell, J.C. & Livert, D.E. (1992). Confidence in the Practice of Clinical Psychology. *Professional Psychology: Research and Practice, 23(5),* 362-368.

Goldberg, L.R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology, 23,* 25-33.

Golden, M. (1964) Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology. 28(5),* 440-446.

Goodyear, R.K. (1997). Psychological Expertise and the Role of Individual Differences: An Exploration of Issues. *Educational Psychological Review, 9(3),* 251-265.

Goodwin, D.W. & Guze, S.B. (1996). *Psychiatric Diagnosis, Fifth Edition.* New York: Oxford University Press.

Groen, G. & Patel, V. (1985). Medical problem-solving: Some questionable assumptions. *Medical Education, 19,* 95-100.

Grove, W.M. (1991). When is a Diagnosis Worth Making? A Statistical Comparison of Two Prediction Strategies. *Psychological Report, 68,* 3-17.

Grueneich, R (1992). The borderline personality disorder diagnosis: Reliability, diagnostic efficiency, and covariation with other personality disorder diagnoses. *Journal of Personality Disorders.6(3),* 197-212.

Hamstra, B. (1994). *How Therapists Diagnose: Seeing through the Psychiatric Eye.* New York: St. Martin's Griffin.

Hardiman, P.T., Dufresne, R., & Mestre, J.P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition, 17(5),*627-638.

Hassebrock, F. & Prietula, M.J. (1992). A Protocol-Based Coding Scheme for the Analysis of Medical Reasoning. *International Journal of Man-Machine Studies, 37,* 613-652.

Hempel, C.G. (1965). Fundamentals of taxonomy. In C.G. Hempel (Ed.), *Aspects of scientific explanation.*(pp. 137-154). New York, NY: Free Press.

Herl, H.E, O'Neil, H.F. Jr; Chung, G.K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior. 15(3-4),* 315-333.

Holt, R.R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal & Social Psychology. 56,* 1-12.

Holzer, C.E., Nguyen, H.T., & Hirschfeld, R.M.A. (1996). Reliability of diagnosis in mood disorders. *The Psychiatric Clinics of North America, 19(1),* 73-84.

Illich, I. (1975). *Medical Nemesis: The Exploration of Health.* London, UK: Ebenezer Baylis and Son limited.

Ivey, A.E. & Ivey, M. (1999). Toward a developmental diagnostic and statistical manual: The vitality of a contextual framework.. *Journal of Counseling & Development. 77(4),* 484-490.

Jaffe, M. (1998). *Adolescence.* New York, NY: Wiley & Sons Publishing.

Johnson, E.J. (1988). Expertise and decision under uncertainty: Performance and process. In M.T.H. Chi, R.Glaser & M.J. Farr (Eds.), *The nature of expertise* (pp. 209-228). Hillsdale, NJ: Erlbaum.

Kassirer, J.P. (1989). Diagnostic Reasoning. *Annals of Internal Medicine, 110,* 893-900.

Kassirer, J.P. & Gorry, A. (1978). Clinical Problem Solving: A Behavioral Analysis. *Annals of Internal Medicine, 89,* 245-255.

Kassirer, J.P., Kuipers, B.J. & Gorry, G.A. (1982). Toward a Theory of Clinical Expertise. *The American Journal of Medicine, 73,* 251-258.

Kaufman, D. R. & Patel, V. L. (1991). Problem Solving in the Clinical Interview: A Cognitive Analysis of the Performance of Physicians, Residents and Students. *Teaching & Learning in Medicine: An International Journal, 3,* 6-14.

Kendell, R.E. (1973). Psychiatric Diagnosis: A Study of How They are Made. *British Journal of Psychiatry, 122,* 437-445.

Kendell, R.E. (1982). The Choice of Diagnostic Criteria for Biological Research. *Archives of General Psychiatry, 39,* 1334-1339.

Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook, Third Edition.* Englewood Cliffs, New Jersey: Prentice Hall, Inc.

Keravnou, E.T. & Johnson, L. (1989). Towards a Generalized Model of Diagnostic Behaviour. *Knowledge-Based Systems, 2(3),* 165-177.

Kingsbury, S.J. (1987). Cognitive Differences Between Clinical Psychologists and Psychiatrists. *American Psychologist, 42(2),* 152-156.

Klein, G.A. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision-making. In G.A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), *Decision-making in Action: Models and Methods,* (p. 138-147). Norwood, New Jersey: Ablex Publishing Corporation.

Klein, G.A. & Calderwood, R. (1991). Decision Models: Some Lessons From the Field. *IEEE Transactions on Systems, Man, and Cybernetics, 21(5),* 1018-1028.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107(3),* 296-310.

Kolodner, J.L. (1984). Towards an understanding of the role of experience in the evolution from novice to expert. *Developments in Expert Systems,* 95-116.

Kolodner, J.L. & Kolodner, R.M. (1987). Using Experience in Clinical Problem Solving: Introduction and Framework. *IEEE Transactions on Systems, Man, and Cybernetics, 17(3),* 420-431.

Kuipers, B. & Kassirer, J.P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8(4),* 363-385.

Kuipers, B., Moskowitz, A.J. & Kassiper, J.P. (1988). Critical Decisions under Uncertainty: Representation and Structure. *Cognitive Science, 12,* 177-210.

Kushniruk, A.W., Patel, V.L. & Fleizer, D.M. (1995). Complex decision-making in providing surgical intensive care. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society,* 287-292.

Meehl, P.E. (1999) Clarifications about taxometric method. *Applied & Preventive Psychology, 8(3),* 165-174.

Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., & Reed, G.M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56(2),* 128-165.

Montgomery, H. & Svenson, O. (1989). *Process and Structure in Human Decision-making.* New York, NY: John Wiley & Sons.

Morrison, J. (1995). *The First interview: Revised for DSM-IV.* New York: The Guilford Press.

Moskowitz, A.J. & Kassirer, J.P. (1988). Critical Decisions under Uncertainty: Representation and Structure. *Cognitive Science, 12,* 177-210.

Moskowitz, A.J., Kuipers, B.J.& Kassirer, J.P. (1988). Dealing with Uncertainty, Risks, and Tradeoffs in Clinical Decisions. *Annals of Internal Medicine, 108,* 435-449.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Nurcombe, B. & Gallagher, R.M. (1986). *The Clinical Process in Psychiatry: Diagnosis and Management planning.* New York: Cambridge University Press.

O'Byrne, K.R. & Goodyear, R.K. (1997). Client Assessment by Novice and Expert Psychologists: A Comparison of Strategies. *Educational Psychology Review, 9(3),* 267-279.

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. & Patel, Y.C. (1997). Cognitive Aspects of Clinical Performance During Patient Workup: The Role of Medical Expertise. *Advances in Health Sciences Education, 2,* 95-114.

Pennington, N. & Hastie, R. (1993). A Theory of Explanation-Based Decision-making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision-making in Action: Models and Methods,* (p. 188-201). Norwood, New Jersey: Ablex Publishing Corporation.

Phelps, R., Eisman, E.J. & Kohout, J. (1998). Psychological practice and managed care: Results of the CAPP practitioner survey. *Professional Psychology: Research and Practice, 29,* 31-36.

Posner, M.I. (1988). Introduction: What Is It To Be An Expert? In Chi, M.T.H., Glaser, R., & Farr, M.J., (Eds.). *The Nature of Expertise,* (p. xxix-xxxvi.). Lawrence Erlbaum Associates Inc.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Reading, MA: Addison Wesley.

Rasmussen, J. (1993). Deciding and Doing: Decision-making in Natural Contexts. In G.A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), *Decision-making in Action: Models and Methods,* (p. 159-171). Norwood, New Jersey: Ablex Publishing Corporation.

Reif, F., & Allen, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. *Cognition and Instruction, 9(1),* 1-44.

Robinson, F.B. (1964). *Introduction to Stuttering.* New Jersey: Prentice-Hall.

Rock, D.L., Bransford, J.D., Maisto, S.S. & Morey, L.,(1987). The Study of Clinical Judgment: An Ecological Approach. *Clinical Psychology Review, 7,* 645-661.

Salovey, P. & Turk, D.C. (1991). Clinical Judgment and Decision-making. In C.R. Snyder & D.R. Forsyth (Eds.), *Handbook of Social and Clinical Psychology The Health Perspective,* (p. 416-437). New York : Pergamon Press.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin. 66(3),* 178-200

Schmidt, H.G., & Boshuizen, H.P. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5(3),* 205-221.

Schmidt, H.G., Norman, G.R., & Boshuizen, P.A. (1990). A Cognitive Perspective On Medical Expertise: Theory and Implications. *Academic Medicine, 65(10),* 611-621.

Schneider, W. (1996). Acquiring expertise: Determinants of exceptional performance. In K.A. Heller & F.J. Moenks (Eds.), *International Handbook of Research and Development of Giftedness and Talent,* (pp. 311-324). Elmsford, NY, US: Pergamon Press, Inc.

Schraagen, J.M., Chipman, S.F., & Shalin, V.L. (2000). *Cognitive Task Analysis.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Shanteau, J. (1988). Psychological Characteristics and Strategies of Expert Decision Makers. *Acta Psychologica, 68,* 203-215.

Shanteau, J. (1992). Competence in Experts: The Role of Task Characteristics. *Organizational Behavioral and Human Decision Processes, 53,* 252-266.

Shanteau, J. (1992). How Much Information Does an Expert Use? Is It Relevant? *Acta Psychologica, 81,* 75-86.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129-138.

Simon, H. (1973). The Structure of III Structured Problems. *Artificial Intelligence, 4,* 181-201.

Slovic, P. Lichtenstein, S. & Fischoff, B. (1988). Decision-making. In R. Atkinson, R. Herstein, G. Lindzey & D. Luce (Eds.) *Steven's Handbook of Experimental Psychology, Volume 2: Learning and Cognition* (pp. 673-738). New York, NY: John Wiley & Sons.

Spitzer, R.L., Forman, J.B., & Nee, J. (1979). DSM-III field trials: I. Initial interrater diagnostic reliability. *American Journal of Psychiatry, 136(6),* 815-817.

Stack Sullivan, H. (1954). *The Psychiatric Interview.* New York: Norton.

Strupp, H.H. (1960). *Psychotherapists in Action.* New York: Grune & Stratton.

Swartz, S. (1991). Clinical Decision-making. In P.R. Martin (Ed.), *Handbook of Behavior Therapy and Psychological Science: An Integrative Approach,* (p. 196-216). New York, NY: Pergamon Press.

Tabachnick, B.G. & Fidell, L.S. (1996). *Using Multivariate Statistics, Third Edition.* New York: HarperCollins College Publishers.

Trotter, R. J. (1986). The mystery of mastery. *Psychology Today, 20(7),* 32-38.

Turner, S.M. & Hersen, M. (1985). The Interviewing Process. In M. Hersen & S.M. Turner (Eds.), *Diagnostic Interviewing,* (pp.3-21). New York: Plenum Press.

Tversky, A. & Kahneman, D (1974). Judgment under uncertainty: Biases in judgment reveal some heuristics of thinking under uncertainty. *Science, 185,* 1124-1131.

Tyrer, P. & Steinberg, D. (1998). *Models for Mental Disorder: Conceptual Models in Psychiatry, Third Edition.* Toronto: John Wiley & Sons.

Van Dijk, T.A. & Kintsh, W. (1983). *Strategies of Discourse Comprehension.* New York: Academic Press.

Veltum, L.G. & Miltenberger, R.G.(1989). Evaluation of a Self-Instructional Package for Training Initial Assessment Interviewing Skills. *Behavioral Assessment 11, 165-177.*

Vicente, K.J. & Wang, J.H. (1998). An Ecological Theory of Expertise Effects in Memory Recall. *Psychological Review, 105(1),* 33-57.

Wheelis, J. & Gunderson, J.G. (1998). A Little Cream and Sugar: Psychotherapy With a Borderline Patient. *American Journal of Psychiatry, 155(1),* 114-122.

Wiggins, J.S. (1973). *Personality and Prediction: Principles of personality assessment.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Wolf, Jr., J.A. (1977). *Collecting Data from Patients.* Baltimore: University Park Press.

Wright, H.J., Stanley, I.M. & Webster, J. (1983). The Assessment of Cognitive Abilities in Clinical Medicine. *Medical Education, 17,* 31-38.

Zimmerman, M. (1994). *Interview Guide for Evaluating DSM-IV Psychiatric Disorders and the Mental Status Exam.* East Greenwich, RI: Psych Products Press.

Zubin, J. (1978). Research in Clinical Diagnosis. In B.B. Wolman (Ed.), *Clinical Diagnosis of Mental Disorders: A Handbook,* (p. 3-14). New York: Plenum Press.

Diagnostic Criteria for Borderline Personality Disorder
(American Psychiatric Association, 1994, p.654)

Diagnostic criteria for 301.83 Borderline Personality Disorder

A pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

(1) frantic efforts to avoid real or imagined abandonment. **Note:** Do not include suicidal or self-mutilating behavior covered in Criteria 5.

(2) A pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation.

(3) Identity disturbance: markedly and persistently unstable self-image or sense of self.

(4) Impulsivity in at least two areas that are potentially self-damaging (e.g., spending, sex, substance abuse, reckless driving, binge eating). **Note:** Do not include suicidal or self-mutilating behavior covered in Criterion 5.

(5) Recurrent suicidal behavior, gestures or threats, or self-mutilating behavior.

(6) Affective instability due to a marked reactivity of mood (e.g., intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely lasting more than a few days).

(7) Chronic feelings of emptiness

(8) Inappropriate, intense anger, or difficulty controlling anger (e.g., frequent displays of temper, constant anger, recurrent physical fights).

(9) Transient, stress-related paranoid ideation or severe dissociative symptoms

The patient portrayed in this study met the diagnostic criteria for Borderline Personality Disorder. Endorsed criteria include (2), (4), (5), (6), (7), (8).

APPENDIX 2

Clinical Knowledge Measures

and

Background Information

Psychometric Test Familiarity

How familiar are you with the following tests, inventories and batteries?

1 2 3 4 5 6 7
never heard of it — administer & interpret with ease

TEST	RATING
Beck Depression Inventory	
Bender Gestalt	
Draw-A-Person/House-Tree-Person	
Millon Clinical Multi-Axial Inventory (MCMI)	
Minnesota Multiphasic Personality Inventory (MMPI)	
Personality Assessment Inventory	
Rorschach Inkblot Test	
Thematic Apperception Test (TAT)	

Clinical Specialties & Theoretical Orientation Checklist

How would you best describe the **type of therapy interventions and theoretical orientations** you use for **most** of your clients? (You may choose more than one)

ASSESSMENT

- ☐ Adult Assessment
- ☐ Child Assessment
- ☐ Forensic Assessment
- ☐ Neuropsychological Assessment
- ☐ Psychiatric Assessment

- ☐ Behavioral
- ☐ Cognitive
- ☐ Cognitive-Behavioral
- ☐ Eclectic
- ☐ Existential
- ☐ Family/Marital Therapy
- ☐ Humanistic
- ☐ Play Therapy
- ☐ Psychoanalytic
- ☐ Psychodynamic
- ☐ OTHER:

MODALITIES you use most often in therapy:

- ☐ Assessment
- ☐ Child/Adolescent
- ☐ Family
- ☐ Group
- ☐ Individual
- ☐ Marital/Couples
- ☐ OTHER:

Please check those **disorders** or **services** you SPECIALIZE in:

- ☐ ADHD/ADD
- ☐ Adjustment Disorders
- ☐ Alcohol or Substance Abuse/Dependence
- ☐ Anxiety Disorders, including Panic Attacks
- ☐ Bipolar/Mania
- ☐ Depression/Dysthymia
- ☐ Forensic/Court Evaluations
- ☐ Gay/Lesbian Issues

- ☐ Men's Issues
- ☐ Women's Issues
- ☐ Neurological Disorders
- ☐ Personality Disorders
- ☐ Psychological Testing
- ☐ Schizophrenia/Delusional Disorders
- ☐ Sexual Disorders
- ☐ OTHER:

Note: The categories used in this clinical checklist were adapted from the APPIC (The Association of Psychology Postdoctoral & Internship Centers) website at www.appic.org.

Semi-Structured Interview

(administered by researcher)

Semi-Structured Interview

1- Describe your educational background and training. Any special awards or distinctions?

2- Describe your clinical activities. Private practice, hospital, clinic settings?

3- How many hours of clinical work do you do in a typical week?

4- Since your internship, how many years have you been practicing full-time?

5- When you meet a patient/client for the first time, how do you usually structure the encounter? Do you have a standard assessment battery? What does it consist of?

6- What percentage of the time do you use DSM-IV diagnostic criteria?

7- What is your age?

APPENDIX 3

Multiaxial Evaluation Report Form

Multiaxial Evaluation Report Form
(American Psychiatric Association, 1994)

AXIS I: Clinical Disorders
Other conditions that may be a focus of clinical attention

_____

_____

AXIS II: Personality Disorders
Mental Retardation

_____

_____

AXIS III: General Medical Conditions

_____

_____

AXIS IV: Psychosocial and Environmental Problems

- ❒ Problems with primary support group. Specify:
  _____
- ❒ Problems related to the social environment. Specify:
  _____
- ❒ Educational problems Specify:
  _____
- ❒ Occupational problems Specify:
  _____
- ❒ Housing problems Specify:
  _____
- ❒ Economic problems Specify:
  _____
- ❒ Problems with access to healthcare services. Specify:
  _____
- ❒ Problems related to interaction with the legal system/crime. Specify:
  _____
- ❒ Other psychosocial and environmental problems. Specify:
  _____

AXIS V: Global Assessment of Functioning Scale
Score: _____

APPENDIX 4

Testing Material: WAIS, MMPI & Rorschach

Referral Note:

Name: Tara

Age: 28

Born: Montreal

Language: Bilingual; French and English

Occupation: Registered Nurse (pediatrics)

Complaint: Stutters; anxious when giving presentations; marital discord

Diagnostic work-up requested

WAIS-R RECORD FORM

WECHSLER ADULT INTELLIGENCE SCALE—REVISED

NAME _____

ADDRESS _____

SEX _____ AGE _____ RACE _____ MARITAL STATUS _____

OCCUPATION _____ EDUCATION _____

PLACE OF TESTING _____ TESTED BY _____

TABLE OF SCALED SCORE EQUIVALENTS*

Scaled Score	RAW SCORE: VERBAL TESTS						RAW SCORE: PERFORMANCE TESTS					Scaled Score
	Information	Digit Span	Vocabulary	Arithmetic	Comprehension	Similarities	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Digit Symbol	
19	—	28	70	—	32	—	—	—	51	—	93	19
18	29	27	69	—	31	28	—	—	—	41	91-92	18
17	—	26	68	19	—	—	20	20	50	—	89-90	17
16	28	25	66-67	—	30	27	—	—	49	40	84-88	16
15	27	24	65	18	29	26	—	19	47-48	39	79-83	15
14	26	22-23	63-64	17	27-28	25	19	—	44-46	38	75-78	14
13	25	20-21	60-62	16	26	24	—	18	42-43	37	70-74	13
12	23-24	18-19	55-59	15	25	23	18	17	38-41	35-36	66-69	12
11	22	17	52-54	13-14	23-24	22	17	15-16	35-37	34	62-65	11
10	19-21	15-16	47-51	12	21-22	20-21	16	14	31-34	32-33	57-61	10
9	17-18	14	43-46	11	19-20	18-19	15	13	27-30	30-31	53-56	9
8	15-16	12-13	37-42	10	17-18	16-17	14	11-12	23-26	28-29	48-52	8
7	13-14	11	29-36	8-9	14-16	14-15	13	8-10	20-22	24-27	44-47	7
6	9-12	9-10	20-28	6-7	11-13	11-13	11-12	5-7	14-19	21-23	37-43	6
5	6-8	8	14-19	5	8-10	7-10	8-10	3-4	8-13	16-20	30-36	5
4	5	7	11-13	4	6-7	5-6	5-7	2	3-7	13-15	23-29	4
3	4	6	9-10	3	4-5	2-4	3-4	—	2	9-12	16-22	3
2	3	3-5	6-8	1-2	2-3	1	2	1	1	6-8	8-15	2
1	0-2	0-2	0-5	0	0-1	0	0-1	0	0	0-5	0-7	1

*Clinicians who wish to draw a profile may do so by locating the subject's raw scores on the table above and drawing a line to connect them. See Chapter 4 in the Manual for a discussion of the significance of differences between scores on the tests.

	Year	Month	Day
Date Tested			
Date of Birth			
Age			

SUMMARY

	Raw Score	Scaled Score	
VERBAL TESTS			
Information			
Digit Span			
Vocabulary			
Arithmetic			
Comprehension			
Similarities			
Verbal Score			
PERFORMANCE TESTS			
Picture Completion			
Picture Arrangement			
Block Design			
Object Assembly			
Digit Symbol			
Performance Score			

	Sum of Scaled Scores	IQ
VERBAL		
PERFORMANCE		
FULL SCALE		

THE PSYCHOLOGICAL CORPORATION
HARCOURT BRACE JOVANOVICH, INC.

9-991829

MMPI

PERSONALITY INVENTORY
S.R. Hathaway and J.C. McKinley

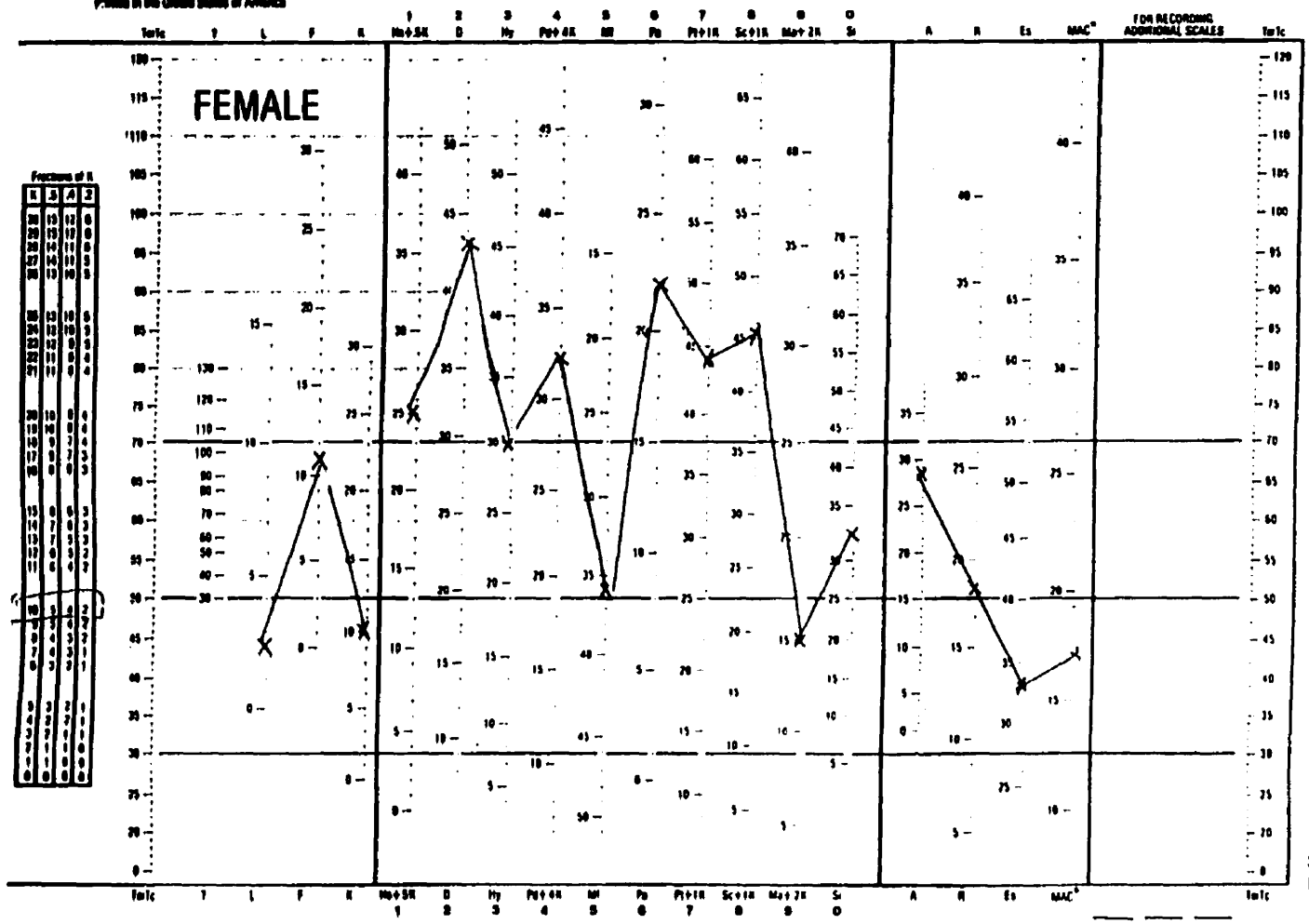
PROFILE

ADDRESS _____

OCCUPATION _____ DATE TESTED / /

EDUCATION _____ AGE ____

MARITAL STATUS _____ REFERRED BY _____



MMPI Code

Scorer's Initials _____

*49 item version

NATIONAL COMPUTER SYSTEMS

27309

Tara's Rorschach

CARD I:
1) skeletal part of human body; hips & coccyx
2) monstrous fly (W) ugly, like its ready for prey

CARD II: "Blood, Oh my God"
1) 2 animals in conflict, fighting, they are hurt (W)
2) a woman's vagina; blood coming out; red

CARD III:
1) 2 people trying to fight out who is going to have it (centre) (W): red signifies blood
2) (upside down): giant roach with claws (W); maybe tarantula
3) (upside down at top): pigs feet
4) (upside down): devil's look; looking down

CARD IV: "Oh my God, they get worse"
1) (laughs, turns card): body being split apart (W) by some weapon
2) upper part interesting; no blood
3) (upside down): 2 heads of dogs

CARD V:
1) bat; animal with horns and wings; almost human
2) 2 heads of crocodiles on sides of the wings
3) (centre): a child's doll; black eye visible
4) cross between a human and object

CARD VI:
1) part of human body, not skeletal; spongy tissue of lungs; does not look good; black (W)
2) snake; only head, little eyes; other part looks like human
3) bottom looks like hooks with eyes or claws with eyes; animal split into 2 with sword; everything is divided into 2

CARD VII:
1) 2 women face to face confronting eachother; identical twins
2) heads of 2 monsters right below their (women's) neck; pigs with horns - monstrous, look angry underneath
3) (bottom/centre): looks like the body of a butterfly without wings stuck between 2 rocks trying to get out
4) (bottom sides): butcher's cleaver

CARD VIII:

1) 2 pink rats or pigs
2) this look like a body to me
3) hands are here - everything split in middle, coccyx, ribs, human: but its also an animal - not clear
4) (upside down): pink animals
5) (upside down): the orange is the brain, then the lungs, and lower down is the pelvis

CARD IX: "I hate this test; this looks terrible"

1) fire, or when blood is washed
2) heads of 2 babies and part of chest, split in 2
3) can't make out green (turns card): (upside down): man standing against a tree, but you have to eliminate colours

CARD X:

1) little baby on top, angels, halo or aura, being lifted up
2) 2 crosses on tips of wings
3) (centre): hips, pelvis, stick something up the vagina, if we consider this female
4) seahorses
5) dolphins
6) (centre): does not look like bones, looks like metal holding things together
7) big splash, everything is exploded (W)
8) witches with tails, both holding animals on their backs and holding torches with green fire
9) dramatic caricatures

Least Liked: Card IV
Most Liked: Card X

All responses were quick; no long delays
No Inquiry