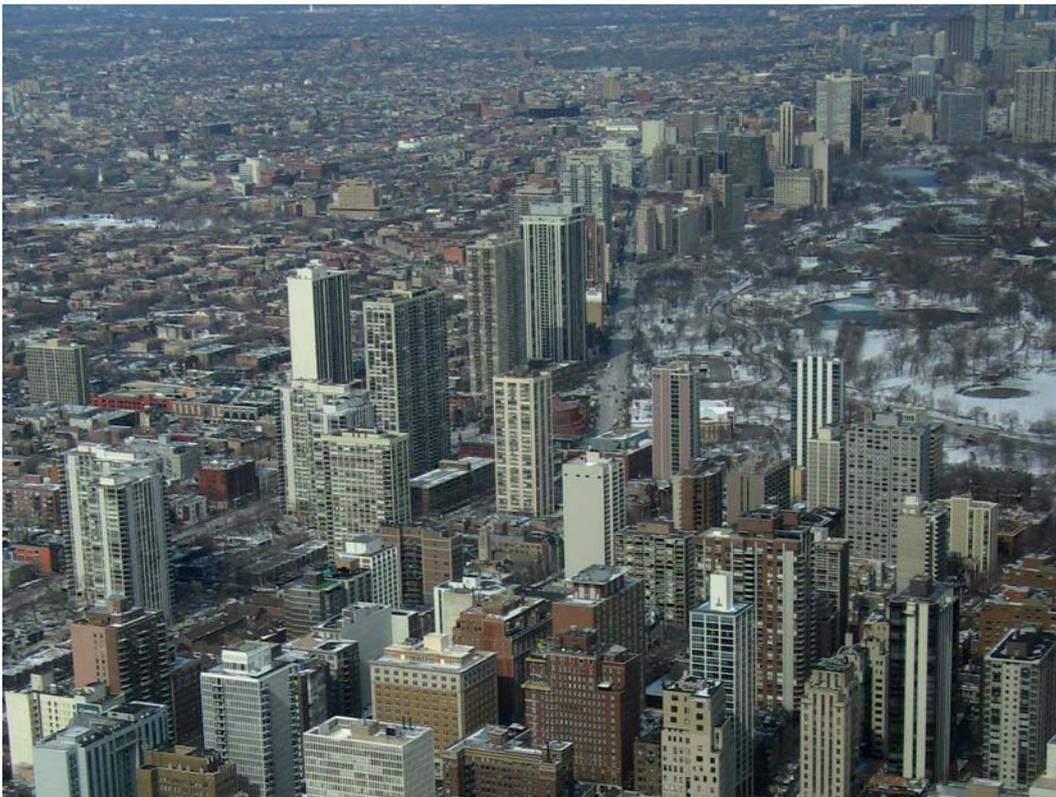Olli-Pekka Santala

Expertise in Using the Rorschach Comprehensive System in Personality Assessment

Olli-Pekka Santala

Expertise in Using the Rorschach Comprehensive System in Personality Assessment

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2009

Expertise in Using the Rorschach Comprehensive System in Personality Assessment

Olli-Pekka Santala

Expertise in Using the Rorschach Comprehensive System in Personality Assessment

UNIVERSITY OF JYVÄSKYLÄ

ABSTRACT

The aim of the present study has been to investigate how differently skilled clinical psychologists use the Rorschach Comprehensive System (RCS) when they assess personality. There are no earlier studies in this area, and therefore the nature of this study is exploratory, aiming at the discovery of new knowledge that can act as a basis for further investigations. The subjects comprised three skill groups (expert, intermediate, and novice groups), each of which included 10 participants. They were to assess the personality of two real patients, using the anamnesis, the WAIS-R profile, and the RCS-materials. The assessments were carried out thinking aloud and the transcribed verbal protocols (610 pages) were used as data. The relevant categories were found using a data-driven approach. They were as follows: items used from the RCS-materials, inferences made from items, integration of information, errors in using the RCS, and content and structure of personality characteristics. The first key result was that experts differed mostly from other groups by making more inferences from RCS-materials, integrating more, making fewer errors and constructing more complex personality characteristics, and skilled use of the RCS in personality assessment is thus primarily seen to be connected to these activities. The second main result was that there were no differences between the intermediate and novice groups, which raises discussion about the development of skills in this area.

Keywords: Expertise, Rorschach Comprehensive System, personality assessment, personality test, think aloud, protocol analysis.

Author's address   Olli-Pekka Santala
Laurinkatu 58 B 24
08100 Lohja
Finland
op.santala@saunalahti.fi

Supervisors   Jarl Wahlström
Department of Psychology
University of Jyväskylä

Carl-Erik Mattlar
Department of Psychology
University of Jyväskylä

Pertti Saariluoma
Department of Computer Science and Information
Systems
University of Jyväskylä

Reviewers   Henny P. A. Boshuizen
Educational Technology Expertise Center
Open Universiteit Nederland

Philip Erdberg
Department of Psychiatry
University of California

Opponent   Leonard Handler
Department of Psychology
University of Tennessee

PREFACE

In 2001-2002 I attended an educational program on the Rorschach Comprehensive System, after which I had the honour and opportunity to have discussions once a month with associate professor Carl-Erik Mattlar. Having earlier studied philosophy and logic, I carried within myself an enthusiasm towards inference processes, and during numerous discussions doctor Mattlar generously helped me to build up an idea for research that could stem from this interest. After we had found the idea of "assessing assessment" but did not have any methodology to carry it out, I called professor Pertti Saariluoma, who had been studying the thinking of expert chess players, and after very briefly summarizing my problem he asked me to come over. At the same time I contacted professor Jarl Wahlström, who was very open for my still very confused ideas.

I owe my deep gratitude to these three men, who made this research possible. Working with them has been a great personal pleasure, and during the years I came to know three very skilled thinkers, each in his own way.

I also wish to thank senior physician and psychoanalyst Jukka Häivä, from the Helsinki University Central Hospital, who offered me substantial help which turned out to be immensely important. I thank PhD Asko Tolvanen and PhD Anna-Liisa Lyyra, both from the University of Jyväskylä, for making the statistical analyses. I'm also grateful to the Department of Psychology, University of Jyväskylä, and Jorvi Hospital/Helsinki University Central Hospital, for financial and material help.

Last but not least, I wish to thank people close to me for their important support. My wife Niina has been gently tolerating my behaviour and opening many new conceptual avenues through discussions, and friends and relatives have been able to tolerate my continuing absence.

Lohja 15.10.2009

Olli-Pekka Santala

FIGURES

TABLES

CONTENTS

1 INTRODUCTION

In health sciences clinical activities comprise two essential elements: instruments that are used to obtain valid and reliable knowledge, and clinicians who use those instruments. In clinical personality assessment this division is made between personality testing and personality assessment. Personality testing stands for the process of using personality tests to obtain information on someone's personality, and personality assessment in turn is the process where obtained information is processed to understand the person and answer referral questions (e.g., Handler & Meyer, 1998). In order to pursue effective clinical practice both of these elements must be empirically well-founded. However, scientifically the situation in personality assessment is poor in the sense that there is a large amount of research regarding some personality tests but very little knowledge concerning the personality assessment process. This lack of research on personality assessment is the major motivation behind this thesis.

In medical diagnostics the situation is different. Medical instruments have naturally been researched a great deal, but during the last 30 years also physicians using those instruments have been studied (e.g., Ericsson, Charness, Feltovich, & Hoffman, 2006). According to that research the thinking of physicians often forms the bottleneck in medical diagnostics (e.g., Gilhooly et al., 1997; Simpson & Gilhooly, 1997). As a result, Gilhooly et al. (1997) made the dramatic conclusion that "[w]hether a correct diagnosis has been made is not always easy to determine short of autopsy" (p. 200). And according to statistics medical error is the eight leading cause of death in the United States – more frequent than deaths in motor vehicle accidents or breast cancer (Kohn, Corrigan, & Donaldson, 1999). Thus, there clearly needs to be research on physicians' thinking processes.

Errors in clinical personality assessment also affect people's lives when they result in incorrect decisions and unnecessary suffering. The fact that there is practically no research on personality assessment is alarming considering the fact that some personality tests are very widely used around the world. According to surveys, in the United States the Rorschach inkblot test has been one of the most used methods since 1940s (e.g., Lubin, Larsen & Matarazzo, 1984). In

Finland, 84% of all Finnish psychologists use tests, testing about 232 500 persons in one year (Psykologi [Psychologist], 8, 2008). In Finland too one of the most used personality tests is the Rorschach.

Although there is scarcity of knowledge on personality assessment, there exists a respectable amount of research behind some tests. Good examples are the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940, 1943) and the Rorschach inkblot test (Rorschach, 1921/1942). Their later versions, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989) and the Rorschach Comprehensive System (RCS; Exner, 1974, 1993, 2003), are considered by many to be as reliable and valid as good medical methods (Meyer et al., 2001; Society for Personality Assessment, 2005). This comparison between clinical psychology and medical diagnostics is flattering and one may be tempted to conclude that personality assessment is empirically as reliable as medical diagnostics. This conclusion however is ill-founded, since it is about the instruments only and not about the way information is processed during personality assessment. Since there are no studies of clinicians, there really is no knowledge about what happens when they practice clinical personality assessment. This lack of research has resulted in a situation where there exists a large mental health field which is not empirically grounded in the sense that no one knows how clinical psychologists actually execute personality assessment. If studies in medical diagnostics apply to personality assessment, then the information processing in personality assessment is dependent on thinking in such a magnitude, that without scientific research not even excellent psychometric properties of psychological tests can rescue assessment from the possible dangers of flawed thinking.

These things give a clear rationale for studying the personality assessment process. There have been several reasons to choose the RCS for this study. First, as already mentioned, the RCS is very widely used. Second, it seems that there simply are no studies of how the information it produces is used in personality assessment. Third, there is a great deal of research in terms of its empirical properties. The Variables it comprises have a good to excellent coding reliability (Exner, 2003), and their validity has proven to be as good as that of the best psychological tests or good medical methods (Meyer et al., 2001; Society for Personality Assessment, 2005). Fourth, by choosing an instrument with good empirical properties and studying how it is used may in future enable still more advanced personality assessment. Fifth, an additional reason to choose the RCS is that it is a complicated method. With roughly one hundred Variables it provides so much information that it presents the clinician with a challenge that goes way beyond measuring blood pressure. How the clinicians master this challenge is a good question.

But why, then, study expertise in the use of a method when most of the clinicians are not experts? Shortly, expertise research offers an efficient way to explore this area. In this paradigm differently experienced clinicians are compared when they perform some task. This gives knowledge about what it takes

to use an instrument well, and at the same time it informs us about what aspects of using it are most demanding.

Thus, the key question of this dissertation is how experts versus less experienced clinicians use the RCS while executing clinical personality assessment. This topic is directly relevant in terms of the quality of those mental health services where patients are being assessed for diagnostic or treatment planning purposes. It is also important in terms of implications it has on teaching the use of the RCS and being able to skillfully assess personality. This subject has been approached by having participants from three different experience groups assess the personality of two actual patients. To execute this, the participants were given patients' anamnestic information, their WAIS-R profiles, and the RCS-materials, and about 45 minutes time for each patient. Participants were to think aloud during the process, and these think-aloud protocols were used as data. This resulted in 610 pages of transcribed data, which was analyzed using both quantitative and qualitative methods.

The rest of this introduction focuses on the main themes of this thesis: personality assessment, personality tests, the Rorschach Comprehensive System, and expertise. The design of the study is described in Chapter 2, and the results in Chapter 3. Chapter 4 discusses the results, their relationships to earlier research, methodological issues, and implications for teaching and future research.

1.1 Personality Assessment

Personality assessment is about a hundred years old practice, depending slightly on how it is understood. One essential landmark in the late 19th century was the emergence of psychoanalysis and the creation of first psychoanalytic models of personality (Breuer & Freud, 1895/1960; Freud, 1899/1955). Another landmark was the creation of the first psychological laboratory in Leipzig in 1879 by Wundt (Nicolas & Ferrand, 1999). In the United States, the beginning of clinical psychology can be dated to Wundt's student Cattell's works in the late 19th century (Cattell, 1928) or to the creation of a psychology clinic at the University of Pennsylvania in 1896 by Cattell's first student Witmer (1907).

In the beginning psychological testing did not aim to obtain information about personality, but rather about phenomena such as intelligence and achievement. With regard to personality tests, Jung's (1910) word association method is seen as the first modern personality test since it relied on standardized administration and used reference data in interpretation (Smith & Archer, 2008). Following that, during World War I, Woodworth created a 116-item Personal Data Sheet to screen army recruits for emotional disorders (Woodworth, 1919, 1920), and Pressey developed the Cross-Out Test to study intelligence and emotional interests with survey scales (Pressey, 1918; Pressey & Pressey 1919). During the last 90 years personality testing and personality assessment have

developed into a large field in e.g. medical, neuropsychological, industrial, organizational and forensic settings (Fernandez-Ballesteros et al., 2001; Krishnamurty et al., 2004).

There seems to be two ways to approach or define personality assessment. The first one defines its contents by contrasting it with personality testing. Personality testing can be understood to mean the process of using personality tests to obtain information of someone's personality, whereas personality assessment is understood as the interpretation, integration and processing of the information obtained through testing (e.g., Archer & Smith, 2008; Handler & Meyer, 1998; Weiner & Greene, 2008). Also psychological assessment more generally has been defined in this way (Society for Personality Assessment, 2006; Ganellen, 1996a; Groth-Marnat, 1999; Fernandez-Ballesteros et al., 2001; Krishnamurty et al., 2004).

The second approach gives personality assessment a wider definition. With this respect a good example was given by a workgroup that evaluated essential factors in competent assessment skills (Krishnamurty et al., 2004). They came to list six factors that define personality assessment: problem identification and case conceptualization; understanding and selecting appropriate assessment methods; effective application of assessment procedures; systematic data gathering; information integration, inference, and analysis; communication of findings and development of recommendations; and feedback that is understandable, useful, and responsive. Thus, this view covers much more than only the processing of information obtained by testing.

The above definitions give some perspective to the meaning of personality assessment. However, the crucial issue here is not to debate about the definition, but to understand the scientific basis of personality assessment. To reach this a systematic framework was constructed for presenting the existing empirical research on personality assessment that uses the Rorschach or the RCS. The extensive personality assessment literature beginning with the Rorschach's publication (Rorschach, 1921/1942) can be divided into six objects of scientific research, which are presented in Table 1. According to it, research can focus on the following objects: administration; coding; interpretation; integration of test items within a single test; integration of information sources; and prediction or clinical judgment or personality description. Each of these objects includes two possible levels of research. The first level is to describe the scientific research of the method (Rorschach or the RCS) through empirical scientific research. This will be called test-level research. The second level is to describe the scientific research of clinicians using the Rorschach, which is called clinician-level research. For example, to study interpretation on the test-level means to study what different elements in the Rorschach mean, and on the clinician-level it means to study how clinicians interpret them. Similarly, on the test-level integration of information sources means how the Rorschach is related to some other test, and on the clinician-level it means how clinicians integrate these tests. This kind of classification presents a systematic way to approach personality assessment and enables one to understand what the scientific basis of it

really is. The relevant literature base is enormous, and although an effort has been made to thoroughly examine the literature, it is possible that there are studies that have not been detected. These conceptual levels are now described in more detail.

TABLE 1   The Field of Scientific Rorschach-Based Personality Assessment, in Terms of the Object and Level of Research

		Level	
	Object	Test	Clinician
1)	Administration		See Exner, 2003
2)	Coding		See Exner, 2003
3)	Interpretation of test items	Shalit, 1965; etc.	No studies found.
4)	Integration within test	No studies found.	No studies found.
5)	Integration between information sources	Weilbaecher, 2002; Dao, 2008; etc.	Only one study found: Potkay & Ward, 1972
6)	Prediction, clinical judgment, personality description	Morris, 1943; Levi, 1951; Rogers, Knauss, & Hammond, 1951: etc.	Chambers & Hamlin, 1957; Armitage & Pearl, 1957; etc.

Administration. There naturally exists only clinician-level research on administration. Test developers have executed research on relevant empirical questions around test administration, and e.g. Exner (2003) reports research concerning seating and inquiry. The results function in giving guidelines for creating normative views on how clinicians should operate. For example, Exner, Leura and George (1976) found how the beliefs of the clinicians resulted in different kinds of protocols if face-to-face seating was used, but not if a side-by-side seating was used. It should be noted, however, that on the clinician-level there seem to be no studies on how clinicians in reality administer the Rorschach or the RCS, that is, whether they actually follow the guidelines or not and to what extent.

  Coding. With less structured tests the question of coding reliability of different test items is very important, and therefore on the clinician-level the coding has been of considerable scientific interest. E.g., Exner (2003) reports the coding reliability to be from good to excellent for all determinants, and Viglione & Taylor (2003) give analysis on questions around reliability. Using these kinds of studies it has been possible to establish normative rules for coding. Again, it

should be noted that on the clinician-level studies on coding reliability are not about how clinicians code in real life, but about how they are able to code in circumstances where they know that they are supposed to perform well.

Interpretation. The most researched personality tests, such as the RCS, have a great deal of research behind test variables. On the test-level, Exner (2003) reports a large number of empirical studies on the validity of Variables. One of these is the study on the Inanimate Movement Responses by Shalit (1965), who found that during marked stress and circumstances that cause helplessness the number of Inanimate Movement Responses increase significantly. Based on these validity studies on the test-level the RCS includes normative views on how the clinician should interpret them. However, although the size of this literature is vast, at the same time there seem to be no empirical clinician-level studies, that is, studies of how clinicians interpret Variables.

Within-source integration. When assessment proceeds from interpretation on, the test items within a test can be integrated. The Rorschach systematizers have been quite unanimous in that Variables within the Rorschach must be integrated (e.g., Exner 2003; Klopfer, Ainsworth, Klopfer & Holt, 1954; Piotrowski, 1957). However, in spite of these views there seem to be no test-level nor clinician-level research on within-source integration.

Between-source integration. The empirical studies of integrating information sources are quite a recent phenomenon. Concerning studies on the test-level, in one of the latest studies Dao (2008) studied the relationship between the RCS and the MMPI-2 related to assessing psychosis. It was found that when test interaction styles were ignored, there were significant correlations between the PTI-index in the RCS and the Sc and BIZ scales in the MMPI-2. In another study Weilbaecher (2002) found that diagnostic accuracy increased when SCZI and BIZ were combined, which thus showed incremental validity between them. Many other test-level studies have appeared concerning incremental and convergent validity (e.g., Ackerman, Hilsenroth, Clemence, Weatherill & Fowler, 2001; Dana & Cunnigham, 1983; Krishnamurthy, Archer & House, 1996; Meyer, 2000).

Concerning studies on the clinician-level, only one study was found that could be evaluated to be on this level. Potkay and Ward (1972) had clinicians use three information sources (Rorschach plus history; history; best guess) in assessing diagnosis, anxiety and intelligence. The assessment of anxiety and intelligence was more accurate when the Rorschach and history were integrated, but assessment of diagnosis was more accurate with the history alone. However, there are no clinician-level studies on between-source integration using the RCS.

Prediction, clinical judgment, personality description. On the test-level this group of studies measures the Rorschach's ability to predict behavior or outcome. These studies have been executed since the 1940s and they mostly study two things. The first group of these concern either treatment outcome or factors related to it. Many studies have examined the ability of the Rorschach Prognostic Rating Scale to predict outcome (e.g., Cartwright, 1958; Endicott &

Endicott, 1964; Frank, 1993; Meyer & Handler, 1997; Newmark, 1973). Also other Rorschach factors have been studied, such as Form Quality (Elfhag, Rossner, Lindgren, Andersson & Carlsson, 2004; Frank, 1993; Levi, 1951), Anatomy Responses (Levi, 1951), the number of Responses (Gibby, Stotsky, Hiler & Miller, 1954; Kotkov & Meadow, 1953), Color Responses (Kotkov & Meadow, 1953), Human Movement Responses (Rogers & Hammond, 1953), Ego Impairment Index (Stokes, Pogge, Powell-Lunder, Ward, Bilginer & DeLuca, 2003), the Rorschach Oral Dependency Scale (Elfhag et al., 2004), Schizophrenia Index (Elfhag et al., 2004), and Food Contents (Elfhag et al., 2004).

The second group concerns prediction of behavior. For example, there are studies concerning the Rorschach's ability to predict academic performance (Carlson, 1969; Cooper, 1955; McCandless, 1949; Munroe, 1945; Thompson, 1948, 1951), adjustment (Argabrite, 1998; Munroe, 1945; Thompson, 1948; Tuber, 1983), aggressive behavior (Frank, 1994; Hughes, Deville, Chalhoub & Romboletti, 1992; Lester, Kendra, Thisted & Perdue, 1975; Long & Miezitis, 1966; Rader, 1957; Young, 1973), suicidal behavior (Arffa, 1982; Daston & Sakheim, 1960; Farberow, 1974; Frank, 1994; Kendra, 1979; Kestenbaum & Lynch, 1978; Neuringer, 1965; Silberg & Armstrong, 1992; Sulfaro, 2005; Zappalaglio, Pavan & Lusignani, 1986), marital compatibility (Bricklin & Gottlieb, 1961) and problem drinking (Sterling, 1995).

On the clinician-level the prediction studies concern clinicians' ability to predict or judge based on the Rorschach. These studies concern predicting diagnosis or nature of personality (Albert, Fox & Kahn, 1980; Armitage & Pearl, 1957; Chambers & Hamlin, 1957; Gunderson & Kapfer, 1966; Turner, 1966; Vandenbert, Rosenzweig, Moore & Dukay, 1964), intelligence (Armitage, Greenberg, Pearl, Berger & Daston, 1955; Bialick & Hamlin, 1954; Grebstein, 1963; Hammond, Hursch & Todd, 1964; Taulbee, 1954), adjustment (Cummings, 1954; Grant, Ives & Ranzoni, 1952; Newton, 1954), treatment outcome (Davids & Talmadge, 1964; Filmer-Bennett, 1955; Winslow & Rapersand, 1964), and miscellaneous objects (Dana, Cocking & Dana, 1970; Forer, Farberow, Meyer & Tolman, 1952; Golden, 1964; Lisansky, 1956; Powers & Hamlin, 1957).

There are no studies on the clinician-level concerning personality description, but there is a great deal of studies describing on a test-level personality characteristics of different groups of people or people having different kinds of disorders. Recent studies on the personality of medical patients include e.g. Parkinsons's disease (Castelli et al., 2008), anorexia (Molgora, 2007), personality disorders (Huprich, 2006), ADHD (Graeff & Vaz, 2006), idiopathic hyperidrosis (Mellano, Salcuni, Zennaro, Lis & Sedona, 2001), Munchausen by Proxy Syndrome (Solomon, 2000), schizophrenia (Friedman, 2000), depression (Dreiss, 2000), dementia (Muzio & Luperto, 1999), alcohol addiction (Bergman, Haver, Bergman, Dahlgren & Nielsen, 1998), migraine (Pisani, Colangeli & Popolla, 1996), agoraphobia (de Ruiter & Cohen, 1993), multiple personality and dissociative disorders (Lincoln, 1993), paraplegia (Mattlar et al., 1993), and learning disabilities (Acklin, 1990).

Overall, this category probably includes the largest number of clinician-level studies. Most of these have been executed during the 1950's and 1960's, and therefore there are no studies using the RCS.

The present situation in research and training. Systematic overview of scientific studies in the Rorschach-based personality assessment thus reveals several important things. First, there exists a great deal of studies on the test-level, but relatively few on clinicians using the RCS. Second, studies on the clinician-level are mostly about administering and coding, but there appears to be almost no clinician-level knowledge on interpretation, within-source integration, or between-source integration of Rorschach with other information sources. Third, although there are clinician-level studies on administering and coding, there are no studies on how clinicians administer the test or code it in real life. Fourth, practically all clinician-level studies on prediction have been done prior to the RCS, and on this level there are no studies of how clinicians use the RCS to describe personality.

In terms of training, surveys highlight the difficult situation in this area. In academic settings, where historically has been a strong emphasis on assessment (Piotrowski & Zalewski, 1993), there now appears to be some decline in the interest towards it (Belter & Piotrowski, 2001). Students do not seem to receive sufficient or qualitatively good education (Childs & Eyde, 2002; Clemence & Handler, 2001; Stedman, Hatch & Schoenfeld, 2001; Watkins, 1991), although at the same time psychological assessment is one of the most important areas for practicing psychologists (Meyer et al., 1998; Watkins, 1991). Thus, while clinical practice and training emphasize assessment on the clinician-level, in present day science this level lacks research. Nevertheless, in spite of these shortcomings, it is possible to execute clinician-level research and extend the scientific basis of using the RCS in personality assessment. This thesis is an effort to that direction.

Further, it goes without saying that practicing clinicians are differently skilled, and therefore their skills can produce differences in their thinking and decisions. In terms of the scientific clinician-level research, the present day situation appreciates the methods to the extent that it loses the clinician. This situation is close to thinking that assessment methods are almost mechanical devises that are similarly pursued by different clinicians. At the same time there exists a vast amount of research from other areas of expertise research showing that skill-related differences do exist and that they do contribute to different results in performance. A good example is medical diagnostics, where there has been research for at least the last 30 years on how the skills and experience of physicians affect diagnostics, or what are the factors that differentiate more and less experienced physicians. For example, Elstein, Shulman and Sprafka (1978) listed five factors they thought are relevant considering medical diagnostic skills: 1) hypotheses are generated early during the diagnostic process; 2) considered hypotheses are limited in number; 3) the most common interpretive error is that of overinterpretation; 4) competencies may be case related; and 5) possession of relevant information and experience are basic in medical diagnos-

tic skills. One result that has been repeatedly found is that skilled physicians execute forward-directed reasoning (e.g., Patel & Groen, 1986a, 1991a), which means that they proceed from the given materials or data towards forming hypotheses that eventually result in a diagnostic solution.

In terms of the validity and reliability of personality tests, it is understandable that there has been a strong effort to develop better methods. But scientists in test-based psychology should not stay away from the clinician-level research. This lack of psychological science of clinician-level personality assessment may be one of the major problems that scientific personality assessment suffers today.

1.2 Personality Tests

1.2.1 A Historical Overview

Psychological tests have been used for about a hundred years. As noted earlier, Jung (1910), Woodworth (1919, 1920) and Pressey (Pressey, 1918; Pressey & Pressey 1919) were the first personality test developers. Woodworth's Personal Data Sheet was the first one-dimensional personality test, meaning that it evaluated one dimension of personality. The first multidimensional personality inventory, Bernreuter Personality Inventory, was published in 1931 (Bernreuter, 1931), and it included scales for neurotic tendency, introversion-extroversion, ascendance-submission, and self-sufficiency. It was the forerunner of many current multidimensional inventories, such as the MMPI, Millon Clinical Multiaxial Inventory (MCMI), Personality Assessment Inventory (PAI), and NEO-PI.

At the same time when first inventories were developed, tests with less structured stimuli, later called "projective" tests, saw daylight. Swiss psychiatrist Rorschach published the Rorschach inkblot test in 1921 (Rorschach, 1921/1942). A forerunner of drawing tests, Draw-A-Man, was published by Goodenough (1926), and Morgan and Murray (Morgan & Murray, 1935; Murray 1943) developed the Thematic Apperception Test (TAT). The concept "projective" was coined by Frank in 1939 (Frank, 1939), and thereafter e.g. the Rorschach was labeled a "projective" method.

During the 1940's and 1950's more sophisticated ways of test construction, scoring and interpretation were developed. During the development of MMPI, Hathaway and McKinley (1940, 1943) were already quite aware of many problems with self-report inventories. Cattell (1949) used factor analysis to develop his Sixteen Personality Item Questionnaire, starting with the huge list of 4500 personality adjectives developed by Allport and Odbert from 18000 English-language words. Among methods using less structure, Machover's (1949) Draw-A-Person test intended to analyze personality through interpreting human figure drawings. A similar test was the House-Tree-Person test by Buck (1948). Using language instead of pictures, the incomplete sentence technique

had been already used earlier as a measure of language ability and intelligence (Rotter & Willerman, 1947), but through the works of Rotter and his colleagues, the Rotter Inclomplete Sentences Blank was developed into the first non-subjective incomplete sentence technique for personality testing (Rotter & Rafferty, 1950; Rotter, Rafferty & Schachtitz, 1949; Rotter & Willerman, 1947).

During 1960s things changed, partly due to the behaviorist movement. Social learning theorists like Mischel (1968) and Peterson (1968) attacked the practice of personality assessment, and humanistic psychologists questioned the morality of assessment, based mainly on the writings of Maslow (1962) and Rogers (1961). At the same time psychologists started to become more inter-ested in conducting psychotherapy as a career. The 40-year period starting in the 1960s is often regarded as an era of shrinkage in the use of personality tests, but many things happened that question this view (Weiner & Greene, 2008). To give some examples, Kinetic Family Drawing Test (Burns & Kaufman, 1970, 1972) and Washington University Sentence Completion Test (Loevinger & Wessler, 1970; Loevinger, Wessler, & Redmore, 1970) were some new modifica-tions of less structured tests. Exner published the Rorschach Comprehensive System (Exner, 1974), which compiled the empirically reliable aspects of earlier work on the Rorschach method. Also among inventories there were several ad-vancements. Based on Murray's framework, Jackson (1974) developed Personal-ity Research Form, measuring 22 aspects of personality. Among the new devel-opments based on the DSM-classifications of psychiatric disorders were Mil-lon's (1977) MCMI, Butcher's MMPI-2 (Butcher et al., 1989) and adolescent ver-sion MMPI-A (Butcher et al., 1992), and Morey's PAI (Morey, 1991). Based on the Five Factor Model, Costa and McCrae (1985, 1992) developed an inventory NEO-PI to measure the respective five factors.

Traditionally, psychological tests have been divided into projective and objective tests, but there is an ongoing replacement of these terms with "per-formance-based" and "self-report", respectively (Meyer et al., 2001; Smith & Archer, 2008). Performance-based tests (earlier projective tests) are less struc-tured tests that function in eliciting response behaviors, which in turn are measured and interpreted. The word "performance" is intended to capture the idea, that what is measured is the behavior a person performs, not his or hers projections. Today, the most studied performance-based tests include the RCS, TAT, sentence completion tests, and drawing tests. Self-reports (earlier objective tests) in turn ask the individual to answer a series of questions about himself or herself. Some of the most studied personality inventories include such meas-ures as the MMPI-2, MMPI-A (adolescent version), PAI, MCMI-III, and NEO-PI (Smith & Archer, 2008).

Surveys done during the 20th century give some view on the attitudes to-wards the Rorschach and personality tests. In their survey Faterson and Klopfer (1945) had 105 responses from academic psychologists, and the results indi-cated mostly positive attitudes towards the Rorschach, a definite place of the method in general psychology, and rather wide use of it in research. Biederman and Cerbus (1971) surveyed changes in Rorschach teaching and found that be-

tween 1961 and 1971 there had been fewer Rorschach courses in universities. According to Exner and Exner (1972), among Rorschach users there was large diversity in training and preferred systems, while 20% did not score the Responses and of those who did 80% changed their primary system in some personal way. Ritzler and del Gaudio (1976) found that in the university programs 81% emphasized the Rorschach but only 24% offered a full year course. Further, experienced teachers rated the Rorschach high but it was not appreciated as a method in research. Ten years later, Ritzler and Alter (1986) found the situation to be quite similar. Hilsenroth and Handler (1995) asked about the attitudes of students, who reported to be unsure of their skills in formulating the Rorschach findings, to lack instruction in using it and to not know the theory well. Thus, concerning the era after the publication of the RCS (Exner, 1974), it seems that teachers appreciated the RCS more than the students were able to handle.

1.2.2 The Rorschach Comprehensive System

The Rorschach has been one of the most used personality assessment instruments since the 1940s. According to Lubin, Larsen and Matarazzo (1984), in a survey from 1959 the Rorschach was the most used psychological test, in 1969 the second most used, and in 1982 the fourth most used psychological test. When an instrument is so widely used, there is a good reason to study how it is used. This section introduces the RCS through its historical roots, general nature, recent developments, and some debates in the literature.

History. The Rorschach inkblot method was originally published by Rorschach in 1921 (Rorschach, 1921/1942). His book, Psychodiagnostics: A diagnostic test based on perception, was originally intended more as a method for identifying mental disorders than for describing personality. After 1921, publications were soon available in German, Spanish, Russian and Japanese (Weiner, 1998). The first English-language article on Rorschach was published by Beck in 1930, where he explained how Rorschach method could be used to measure both intelligence and emotional stability versus instability. After Beck's article, in the United States the Rorschach was quickly developed into five largely incommensurable systems, developed by Beck, Hertz, Rapaport and Schafer, Piotrowski, and Klopfer (Exner, 1969). During the 1960's Exner started to systematically compare these approaches and study their nature, finding large differences between them (Exner, 1969). These included major disagreements in all three important areas of the Rorschach: administration, coding, and interpretation. In addition, at the same time survey data indicated that a substantial percentage of clinicians using the Rorschach were not following any system faithfully, but instead combining them eclectically (Exner & Exner, 1972). In the same survey it was also found that many favored a thematic approach instead of a structural approach. The Rorschach world was therefore rather disorganized and assessment with the Rorschach, if not unreliable, at least too subjective to be scientifically appreciated.

Exner took the overwhelming task of integrating the five major approaches into one empirically reliable and valid instrument, which he labeled

the Rorschach Comprehensive System. The name reflected his goal to include empirically well working aspects of all the five previous approaches. It was originally published in 1974 (Exner, 1974), and it has thereafter been continuously developed and revised (Exner, 1978, 1986, 1991, 1993). The latest version was published in 2003 (Exner, 2003), with additional edition on advanced interpretation two years later (Exner & Erdberg, 2005). Born in 1928, Exner died in February 20, 2006.

Today the RCS has clear standards for administration and coding, with good to excellent coding reliability (Exner, 2003). According to PsychInfo search (All text), in December 2008 there were about 9400 studies about the Rorschach or using it. The validity of the RCS is as high as the validity of the MMPI-2 and Wechsler tests (Hiller, Rosenthal, Bornstein, Berry & Brunell-Neuleib, 1999; Meyer et al., 2001). Exner (1993) published normative samples for 700 nonpatient adults, 1390 children and adolescents between ages 5-16, 320 hospitalized schizophrenics, 315 inpatient depressives, 440 diagnostically diverse outpatients, and 180 outpatients with character disorders. Ten years later Exner (2003) published a new sample with 600 nonpatient adults, and in the beginning of 2007 the Journal of Personality Assessment published an additional number summarizing the latest reference data around the world (Shaffer, Erdberg & Meyer, 2007).

The nature and use of the RCS. In practice, the process of using the RCS is preceded with creating a workable relationship with the client. After that the ten inkblots are shown to the client, who is instructed to tell what they might be, what in the blots the percepts were based on, and where in the blots the percepts were located. After that the verbalizations are coded and the information is interpreted. The guidelines for interpreting the RCS are well reported in textbooks (e.g. Exner, 2003). Most of the interpretation is based on the numerical information, the so called Structural Summary (Appendices 1 and 2), which consists of 8 Clusters of Variables (Control and Stress Tolerance, Situation-Related Stress, Affective Features, Information Processing, Mediation, Ideation, Self-Perception, Interpersonal Perception). Each of these represents some specific personality area with a cluster of statistically obtained Variables. The Structural Summary contains about 100 numerical information points in total.

Interpretation begins by scrutinizing Cluster Search Order (Appendices 1 and 2), which indicates in each individual case in what order the Clusters should be interpreted. The Cluster Search Order is based on Key Variables (Exner, 2003, pp. 228-229), of which the first positive Variable determines the order of interpretation. For example, in Case A in Appendix 1, the first positive Key Variable is Lambda, since it's value is greater than 0.99. The respective Cluster Search Order starts therefore with Processing, proceeds to Mediation and Ideation, and then to Controls, Affect, Self-Perception, and Interpersonal Perception (p. 228). Within each Cluster, Variables are reviewed in a predetermined order (p. 225), called steps. RCS includes totally 78 interpretational steps and 5 other Variables to be taken into consideration. In addition, there is much more information that can be obtained from the RCS. There are the Indexes and the Con-

stellations Worksheet (what Variable constellations are positive within each Index), Sequence of Scores showing what kind of code sequences are connected to each Response, and patient's verbatim Responses together with Inquiry verbalizations. In the RCS the focus is on interpreting Variables and the interpretation of Responses has a secondary role. Historically, in Rorschach's monograph (1921/1942) the emphasis was on structural data, but later there were different attitudes towards the matter. Powers and Hamlin (1957) found that clinicians tended use Responses more than Variables, which was a similar finding to that of Exner and Exner (1972). Because the interpretation of Responses had a poor reliability and validity, Exner (1974, 1993, 2003) diminished their role.

So far the RCS interpretation process was described as an empirically guided stepwise process, where information is interpreted in a predetermined order. Exner's intention was that this interpretation process is not a set of fragmented interpretations, but right from the beginning of the interpretation process Variables are understood as being related to each other. If these relationships are discarded, "the likelihood of error is substantial because this kind of interpretive strategy ignores additional data that will bear directly on whether the deviation based hypothesis should be accepted, rejected, or modified" (Exner, 2003, p. 219). Further, "[d]isconnected hypotheses inevitably lead to a disconnected picture that, at best, fails to capture the organization of the person, or even worse, presents misleading or distorted conclusions" (p. 219). Exner's view is that every new Variable's interpretation adds to the picture of the individual: "As the overall data are reviewed, findings become additive and gradually the simple hypotheses that were formulated early in the interpretive process become broader and more precise, and ultimately capture something of the uniqueness of the individual who has taken the test" (p. 220). Further, in the same way as the interpreted Variables must be connected to other Variables, the interpretation of Clusters must be connected to other Clusters: "Interpreters should strive for a conceptual integration of findings, that is, findings from each cluster are integrated with findings from the other clusters so that any hypotheses or conclusions are based on the total available information" (p. 230). Exner's reasons for stressing integration are that valid interpretation and assessment necessitate integration and that discrete interpretations are unable to reflect the complexity of personality. Only an integrated picture can reveal something about an integrated entity. Finally, Exner writes about integration on the level of assessment: "The final description requires an intelligent merging of the findings, both positive and negative, so that an overall portrait of a unique person gradually evolves and is cast in a framework that addresses the assessment issues that have been posed" (p. 230). Exner thus writes about integration on three different levels: between Variables, between Clusters, and between information sources. However, he does not give clear guidelines on how to execute these integrations. This does not in any way contaminate assessment based on the RCS, but it gives some idea of how far the standardization of interpretation process has proceeded.

Although the above ideas of Exner (2003) about integration are stated on the clinician-level, it is notable that on this level no empirical studies were found utilizing the RCS. Only one clinician-level study on between-source integration using the Rorschach was found, but this was published prior to the RCS (Potkay & Ward, 1972). In terms of the RCS there exist only descriptions by Rorschach scholars, who seem to be single-minded on the need for integration in assessment. For example, according to Weiner (1993) and Ganellen (1996a) differences between assessment methods bring advantages when they are used to complement each other in assessment. Among test-level studies on integration a subject that has lately been of comparatively large interest has been the integration of the RCS and MMPI-2 (Acklin, 1993; Alpert, 1999; Crawford, 2004; Finn, 1996; Ganellen, 1996a; Lovitt, 1993; Meyer, 1997; Weiner, 1993). Also integration of other methods with the RCS has been of some interest. For example, Dorr (1997; 2008) has illustrated the integration of the RCS and the MCMI-III. Fassino, Amianto, Levi and Rovera (2003) executed an empirical study on integrating the RCS with the Temperament and Character Inventory, and found that they validated each other.

Information produced by the RCS does yet not constitute a meaningful description of personality, which necessitates that the psychologist processes the obtained information. That is, in real life there is no personality assessment without clinician-level. It is where testing ends that assessment begins, and this distinction is crucial for understanding what we know and what we do not know about what traditionally has been called personality assessment. For example, while interpreting Variables a clinician may proceed in three ways. He can interpret them in isolation using Exner's interpretation steps (saying, for example, that if S=8, the patient is angry); he can integrate Variables to reach interpretations (saying for example, that AdjD=+1 does not mean better than normal capacity for control since FM=0); or he can integrate Variables with other information sources (for example, AdjD=+1 does not mean better than normal capacity for control since that is in contradiction with the patient's anamnesis). From a wealth of interpreted information one must select those that are thought to be most relevant, for which there are no strict rules. At its best, it is not only an activity with test Variables, but also includes theories and models of personality and psychopathology. The decisions between these depend on personal skills, preferences, and practical purposes, and there are no strict rules for this either. Finally, a description of personality must be produced, for which there again are no rules. Decisions have to be made concerning the conceptual level where personality characteristics are described. For example, are they close to the experiential level, describing mostly patient's experiences, or are they on an empirical level using the conceptualizations of personality tests, or are they on a level of theoretical language speaking, for example, about object relations or neurotic conflicts, or are they close to psychiatric nosology talking in terms of psychiatric disorders, or are they a combination of some of these? These theoretical clinician-level descriptions lack empirical grounds, and the very aim of this thesis is to provide such information.

Some recent developments. Among the new developments of the RCS, Exner's latest book (Exner, 2003) included four new Variables: XA% and WDA% measuring conventionality in perception, Perceptual-Thinking Index (PTI) indicating potential difficulties in reality testing and ideational clarity, and GHR:PHR signifying adaptive interpersonal functioning. In addition, research has been done that aims at building new Variables into the RCS, which were not yet included in the 2003 version of it. For example, Viglione and Perry (e.g., Perry, McDougall & Viglione, 1995; Perry & Viglione, 1991; Perry, Viglione & Braff, 1992) have developed the Ego Impairment Index that is a general measure of psychopathology. With his colleagues, Perry (Perry & Potterat, 1997; Perry, Potterat, Auslander & Kaplan, 1996) has developed new ways to use the RCS as a neuropsychological instrument. As a result from studies regarding psychopathy, Gacono and Meloy (1994) produced new ways to code representations of aggression, which can be used as indicators of potential psychopathy. In addition to these examples, the RCS has been used in numerous studies on personality structures in terms of different psychopathology. These are just a few examples of ways that Exner's Comprehensive System has been developed on the test-level.

Criticisms and discussions. Although the RCS has facilitated research and developments in numerous different areas, and has been of large interest and use around the world, it has also been criticized in different ways. It seems that all of the discussion is focused on the test-level, that is, on the nature of this method. There are two good reviews on this matter, by Exner (2003) and Mattlar (2005), on which this section is mostly based on.

According to Mattlar (2005), a great deal of criticism towards the RCS is ill-founded in that the writers either ignore scientific data or misunderstood research. In a similar fashion Hibbard (2003) points out how some of the critiques either misinterpret or underreport scientific studies. Further, some of the criticism has been questioning the use of the RCS without any deeper or more elaborated criticism (e.g., Dawes, 1994). Thinking about the criticism towards the RCS for example in Finland during the last 10 years, one cannot but wonder about the great need of some scientists or clinicians to criticize the RCS, and it resembles the criticism often put against psychoanalysis.

However, especially during the 1990's there started to appear criticism which was more founded and which prompted productive dialogues in the journals. This was originated by Nezworski and Wood (1995), who questioned the validation data of Reflection Responses and Egocentricity Index. Also Wood, Nezworski, and Stejkal (1996) questioned findings concerning the interscorer agreement, validation data, and unpublished data concerning the RCS. In their reply, both Meyer (1999) and Viglione (1999) demonstrated the empirically valid nature of the RCS. At the same time Hunsley and Bailey (1999) questioned the clinical value of the RCS. Hiller et al. (1999) replied using meta-analysis to show how the RCS and MMPI-2 had similar validity, although with different criterion variables. Similar findings had been obtained earlier by Ganellen

(1996b), according to whom the RCS, MMPI, and MCMI-II were comparable in terms of their diagnostic efficiency.

In 1999, Shaffer, Erdberg, and Haroian (1999) published findings on 123 nonpatients, and it was noticed how it included notable differences with the latest nonpatient sample of Exner (1993). This was a problem within the Rorschach community, and it prompted a new project for collecting nonpatient data (Exner, 2002). Wood, Nezworski, Garb, and Lilienfeld (2001) however used this as a foundation of further criticism, concluding that Exner's sample would not be representative of nonpatients. However, by now, with the publications of new reference data (e.g., Exner, 2003) this criticism seems ill-founded. Concerning interscorer reliability, Wood et al (1996) suggested it to be poor, but Meyer et al. (2001) obtained findings according to which reliability coefficients were generally excellent. The same finding was published by Janson (1999).

Still another wave of criticism appeared in 2001, when Garb, Wood, Nezworski, Grove, and Stejkal (2001) gave further support for their criticisms in terms of lack of evidence for temporal stability for all Variables in the RCS, inability to use the RCS in making clinical diagnoses, and lack of incremental validity with minority populations. As a response, Rosenthal, Hiller, Bornstein, Berry, and Brunell-Neuleib (2001) produced more data supporting their earlier view on diagnostic efficiency. Another reply was published by Viglione and Hilsenroth (2001), who claimed that much of the criticism was biased or without merit. Further, Meyer and Archer (2001) presented a new meta-analysis showing how the validity of the RCS is comparable to the MMPI and Wechsler tests.

In retrospect, it seems that even when the critiques have been ill-founded, they have functioned to sharpen the empirical basis of the RCS. A good illustration of this is a large review produced by Meyer et al. (2001) who used over 125 meta-analyses and 800 multi-method assessments, and showed how the validity of the best psychological assessment methods is comparable to other clinical measures throughout the health sciences. Some psychological instruments (e.g. the RCS, MMPI, Wechsler tests) were shown to perform as well as a large variety of medical methods, e.g., electrocardiograms, mammography, dental radiographs, Papanicolaou (Pap) smears, positron emission tomography (PET) scans, and serum cholesterol level testing (p. 135). It seems that according to critical scientific research the RCS is empirically sturdy, and it is appropriate to use it in assessment. In the end, it may be that the serious criticism of the late 1990s was at least partly provoked by the bad old reputation of the Rorschach as something of a subjective-projective test. Today, however, the RCS is more properly conceptualized as a cognitive and perceptual problem-solving task (Exner, 2003).

1.3 Expertise

The primary aim of this thesis is to study expertise in the RCS-based personality assessment, and therefore in this section an introduction to expertise research is presented. Expertise can be taken to mean "the characteristics, skills, and knowledge that distinguish experts from novices and less experienced people" (Ericsson, 2006a, p. 3). Expertise research offers a way to approach the nature of human problem-solving, in terms of both design and methodology, both of which have been utilized in this study. In terms of the design, the paradigm of using different skill groups is borrowed from expertise research. In terms of methodology, the method of think-aloud or protocol analysis is also borrowed from expertise research. Additionally, the history of expertise research shows how the earlier results in some specific problem-solving environment is not directly applicable to other environments. This section outlines these themes. It starts with a historical overview, presents some research on expertise in both psychological and medical diagnostics, and finally presents the think-aloud method.

1.3.1 A General Review

Historical account. People have probably always been interested in how one can perform better in certain area, or who to rely on with problems in some specific area. During Middle Ages craftsmen formed guilds, which often obtained a monopoly on providing particular types of handcrafts and services with set quality standards (Epstein, 1991). In most guilds an inspection of best work and masterpieces was required, and some guilds administered special tests to assess the level of performance (Epstein, 1991). The scholar's guild was established in the 12th and 13th century as a "universitas magistribus et pupillorum", guild of masters and students (Krause, 1996). Later, in the first universities, to become a master the advanced student needed to satisfy "a committee of examiners, then publicly defend a thesis, often in the town square and with local grocers and shoemakers asking questions" (p. 10). The universities and doctorate programs formed one centerpiece in seeking expertise in the academic fields. Nowadays we are all more or less familiar with different kinds of schools, university faculties, professional societies, clubs, training programs, scholarships, etc., that have evolved around gaining or guaranteeing expertise in a specific field. Also, much of western history has been written around exceptional expert individuals, who established novelties in their specific fields.

The systematic research of expert performance started in the 20th century. Feltovich, Prietula & Ericsson (2006) describe three main lines in this development. The earliest one used the think-aloud methodology. Participants were instructed to think aloud while solving everyday life problems (e.g., Duncker, 1945) or to think aloud while selecting moves for chess positions (de Groot, 1946/1965). Another line developed from the studies on judgment and decision

making, where researchers compared the judgments of experts to those of statistical models (e.g., Meehl, 1954). Perhaps the most important line, according to Feltovich et al. (2006), was one inspired by computational methods and the information processing model connected to it (Miller, Galanter & Pribram, 1960; Newell & Simon, 1972; Reitman, 1965). Some of the most groundbreaking progress came from using information processing modeling in studies on chess expertise (Chase & Simon, 1973a, 1973b). In addition to these, the development of cognitive psychology made it possible to create much more refined ways to model cognitive phenomena associated with expertise.

Expertise has been studied in numerous areas. For example, The Cambridge Handbook of Expertise and Expert Performance (Ericsson, Charness, Feltovich & Hoffman, 2006) describes expertise research from such diverse areas as medicine and surgery, transportation, software design, professional writing, music, sports, arts, chess, and mathematics. It also analyzes expertise from many different perspectives, such as psychology, education, sociology, neuroanatomical changes, aging, social factors, and creativity.

Expertise and task environment. One basic element of this study is that different phenomena are relevant in different problem-solving environments, and therefore one cannot generalize from one task environment to another. In one of the earliest large scale studies on expert thinking, Thought and choice in chess by de Groot (1946/1965), the focus was on how players selected best next moves and remembered chess positions. He found that expert chess players could memorize chess positions in much more detail than less skilled players. After de Groot's seminal work, studies in chess and physics expertise continued focusing on decision making and memory. It was therefore natural to think that decision making and memory constitute the ingredients differentiating experts from less skilled subjects. However, based on their studies in medical diagnostics, Patel and Groen (1991b) suggested that memory as a measure of physicians' expertise is not valid. They showed how in medical diagnostics experts' recall was significantly less detailed and explicit than novices' (Patel & Groen, 1991b). This was explained by experts focusing attention only on relevant case material, whereas novices processed everything with more or less equal attention. They (Patel & Groen, 1991b) argued that studying memory in medical diagnostic expertise was irrelevant, and that expertise was differentiated by an ability to identify relevant information. This is only one example of how expertise is specific to a specific problem-solving environment, and therefore results in one problem-solving area cannot be generalized to another one.

General properties of expertise. According to Glaser and Chi (1988), in the study of expertise the following general features of expert performance have been noticed: experts' performance is limited to their specific area of expertise; they easily perceive large meaningful units; they are faster and make fewer mistakes; their short term and long term memories are better in their specific task domain; they see and represent problems on deeper levels; they use more time to analyze problems qualitatively; and they have better skills to reflect on themselves. These results are, of course, highly general and neglect the

fact brought up earlier, that different problem-solving environments produce different results.

To look at expertise from a different perspective, what does it take for someone to develop into an expert in some field? First of all, intelligence does not seem to be a necessary requirement, although intelligence and working memory are important during the early stages of learning, and more intelligent people learn more quickly (Hunt, 2006). Second, expertise requires a great deal of interest and motivation in a specific field, since otherwise it is not be possible to invest the huge amount of effort it takes to be an expert (Hunt, 2006). However, even though intelligence does not make an expert, this talent is not non-important and works together with motivation. Talents are channeled by interests, and people are more interested in things they are good at than things they find difficult. Third, the social support provided during learning is extremely important. Gardner's (1993) biographical study of exceptional contributors, such as Einstein and Picasso, stresses how these great masters were able to be single-minded because they were supported by family, friends, and colleagues, often at considerable expense. Another kind of social support is provided by society in the forms of salaries and scholarships. Finally, acquiring expertise requires extensive amount of deliberate practice, which, according to some, takes at least ten years (Richman, Gobet, Staszewski & Simon, 1996). So, to be able to accomplish a very high level of performance takes a great deal of interest, social support, and deliberate practice.

Expertise in medical diagnostics. Among the numerous areas of expertise research, medical diagnostics is the one closest to clinical personality assessment, at least in the sense that they both aim at defining what is wrong with the patient. Studies in medical diagnosis have clearly shown how the accuracy of diagnosis depends on thinking, specifically, the knowledge structures used therein (e.g., Barrows et al., 1978; Elstein, Shulman & Sprafka, 1978; Feltovich et al., 1984). They have also had direct practical influence. For example, early studies on the difference between experts and novices (Barrows, Feightner, Neufeld & Norman, 1978; Elstein, Shulman & Sprafka, 1978) led directly to the creation of new methods of instruction and the creation of problem-based learning (Barrows & Tamblyn, 1980).

Norman, Eva, Brooks, & Hamstra (2006) give a historical outline consisting of three approaches in the research of medical diagnostic expertise. First, in early studies medical expertise was equated to general reasoning skills, but it turned out that the same reasoning strategy appeared in all skill levels. The second approach utilized the idea that the amount of knowledge was essential in expertise. However, results showed that experts often recalled less than less skilled physicians. In the third approach researchers turned to examine the type and organization of knowledge. So far these studies have examined three types of knowledge. Causal knowledge refers to scientific knowledge about causal mechanisms, analytical knowledge to knowledge about signs and symptoms, and experiential knowledge to an accumulated storehouse of prior cases that has been constructed based on experience. It seemed likely that each of these may be used differently.

Therefore, a new question was proposed concerning the conditions of applying different kinds of knowledge in different clinical contexts. The research on the nature of expert knowledge is still very much in progress.

Expertise research has repeatedly shown how experts possess more knowledge in their task environment than less experienced individuals (e.g., Chi, 2006; Horn & Masunaga, 2006). Knowledge has been divided into declarative and procedural knowledge (e.g., McNamara, 1994). Declarative knowledge means propositional knowledge that can be verbalized, whereas procedural knowledge consists of knowledge of how to do things. Declarative knowledge can be further divided into scientific (deep) versus clinical (shallow) knowledge (Chi et al, 1981; Gilhooly, 1996; Patel & Arocha, 2001). In terms of errors in medical diagnostics, cognitive mechanisms are the major factor that contributes to medical errors (Patel, Arocha & Zhang, 2005). They may be for example about declarative or procedural knowledge (Patel, Arocha & Zhang, 2005). The integratedness of knowledge refers to the degree to which concepts and principles are related to one another in different meaningful ways (Chi et al., 1981; Horn & Masunaga, 2006; Lesgold et al., 1988; Patel & Arocha, 2001). The concept of chunk has been used to describe the way experts' knowledge structures are larger and more organized and integrated (Chase & Simon, 1973a, 1973b; de Groot, 1965).

Another difference is that between forward reasoning and backward reasoning (Gilhooly, 1996; Nickerson, 1994; Patel & Groen, 1986a; Patel & Groen, 1991a). In forward reasoning a person works from the given information towards the solution, whereas in backward reasoning the working proceeds from a hypothetical solution to the given information. It has been noticed that experts make greater use of forward reasoning while novices tend to reason backwards (Patel & Groen, 1986a). These two types are different in terms of errors made during problem-solving. Forward reasoning is more error-prone since the inferences cannot be checked, and it thus necessitates a great deal of knowledge. While experts have more knowledge in their domain than novices, experts can use this knowledge to solve problems by using forward reasoning (Gilhooly, 1996). It has been found that experts who use forward reasoning are more successful in making diagnoses, whereas the reasoning processes that result in incorrect diagnoses include both types (Patel & Groen, 1986a).

The concept metacognition refers to knowledge about one's performance (Flavell, 1979). It is what an individual knows about one's cognitive processes. According to the research, experts are better in reflecting and self-monitoring than less experienced individuals (e.g., Chi, Feltovich & Glaser, 1981; Larkin, 1983). Experts are better aware of their strengths and weaknesses, and are thus more able to monitor and evaluate themselves during performance.

How to acquire expertise? The prominent way to approach the practice of acquiring expertise has been developed by Ericsson (Ericsson, 1996; Ericsson, Krampe & Tesch-Römer, 1993). The focus of this approach is on the concept of deliberate practice, developed by Ericsson et al. (1993) while studying expertise in musical performance. Deliberate practice means that "the most effective

learning requires a well-defined task with an appropriate difficulty level for the particular individual, informative feedback, and opportunities for repetition and corrections of errors" (pp. 20-1). It "presents performers with tasks that are initially outside their current realm of reliable performance, yet can be mastered within hours of practice by concentrating on critical aspects and by gradually refining performance through repetitions after feedback" (Ericsson, 2006b, p. 692). Evidence for the role of deliberate practice in the development of expertise has been widely discussed, for example, in relation to professional writing, music performance, sports, chess, exceptional memory, and mathematical calculation, self-regulated learning, successful training in simulators, maintained performance in older experts, and creative activities (Ericsson, 2006b). It seems to be an empirical fact that skilled performance is related to deliberate practice (Ericsson, 1996; Ericsson et al., 2006).

Playful interaction or enjoyable activities are usually not included in deliberate practice – indeed empirical findings suggest that experts usually do not find deliberate practice enjoyable (Ericsson, 1996; Starkes, Deakin, Allard, Hodges & Hayes, 1996). It is also not identical to work, since that consists mostly of repeating routine activities. This is understandable, since deliberate practice consists of identifying weaknesses and choosing practice focus based on those weaknesses. This is followed by intentionally practicing weaknesses in performance with appropriate feedback. The context and amount of time devoted to practice is very important. Deliberate practice is often done in total solitude and/or early in the morning in order to maximize concentration. World-class chess players or music performers habitually practice for about four hours per day, without which the yield of practice starts to decline. It is also important to notice that for chess players practice does not mean playing chess, nor does it mean playing musical pieces for music performers. What it means is that weaknesses in performance are deliberately practiced in order to broaden the scope of skilled performance. Also, it is almost a rule that world class experts practice seven days a week.

1.3.2 Think-Aloud Method

Methodologically, this study utilizes the think-aloud method (protocol analysis), which is a common method in expertise research (e.g., Ericsson et al., 2006; Ericsson & Simon, 1993). There the participants are given stimuli and a problem-solving task which they have to perform thinking aloud, and their verbal reports are used as data. Historically, think-aloud method has been used for almost a hundred years. In the beginning of 20th century subjects' think-aloud protocols consisted of the examiners' notes of verbalizations, since there were no tape recorders available. In terms of validity of verbal data this of course was a problem. Probably the earliest documented analysis of a think-aloud protocol was made by Watson (1920), who used it to illustrate some general characteristics of cognitive process in problem-solving. Duncker (1926) analyzed a number of protocols for over 20 problems, trying to find the mechanisms that generated the solutions, and several studies (e.g., Bulbrook, 1936; Claparéde, 1934; Durkin,

1937) continued Duncker's work. In one of the most groundbreaking studies de Groot (1946/1965) recorded by hand extensive protocols of chess players' thinking while they were selecting moves. With the introduction of tape recorders after World War II the situation changed, for it was then possible to operate without any real time constraints. Later, think-aloud method was developed by Ericsson and Simon (1993) into a fully scientific method with proven validity.

In modern science think-aloud method has been widely used in a variety of different fields such as medical diagnostics, chess playing, physics problem-solving, economics, architectural planning, taxi driving, engineering, accounting, learning disabilities, development of survey questions, validation of multiple-choice questions, user testing of computer products, etc. (Ericsson & Simon, 1993). In medical diagnostics alone it has been used, for example, in the diagnostic tasks of gastroenterology (Coderre, Mandin, Harasym & Fick, 2003), breast pathology (Crowley, Naus, Stewart & Friedman, 2003), treatment decisions (Kushniruk, 2001; Kushniruk, Patel & Fleiszer, 1995), visual microscopic diagnosis (Crowley, Naus & Friedman, 2001), pharmacotherapeutics (Boreham, Mawer & Foster, 2000), endocrine problems (Kushniruk, Patel & Marley, 1998), surgical decision making (Lorenz, Schult & Rothmund, 1990), electrocardiogram interpretation (Gilhooly et al., 1997), X-ray pictures (Lesgold et al., 1988), and congenital heart disease (Feltovich et al., 1984). It has also been used in some other kinds of approaches, investigating, for example, how practical guidelines affect diagnostics (Patel, Arocha, Diermeier, How & Mottur-Pilson, 2001) or what is the nature of user-computer interaction in health care (Kushniruk et al., 1998). Diagnostic behavior between nurses and physicians have been compared (Di Giulio & Crow, 1997; Offredy, 2002), and the medical diagnostic research of nurses is quite well represented (e.g., Cioffi & Markham, 1997; Corcoran-Perry, Narayan & Cochrane, 1999; Fisher & Fonteyn, 1995; Fonteyn, 1997; Fonteyn & Fisher, 1995; Fonteyn & Grobe, 1992; Fonteyn & Grobe, 1994; Fowler, 1997; Greenwood, Sullivan, Spence & McDonald, 2000; Grobe, Drew & Fonteyn, 1991; Lamond, Crow & Chase, 1996; Redden & Wotton, 2001).

For Ericsson and Simon (1993) think-aloud method is both a way to collect data and a way to encode data. However, in an exploratory study like this dissertation it is not meaningful to use protocol analysis to classify data. This kind of use would mean to decide a-priori the classes of analysis and this would contradict the exploratory aim of this study, namely, to discover a data-driven description of phenomena that come up in the personality assessment process. With the lack of prior research in the specific task environment of RCS-based personality assessment there is no prior information as to what kind of classifications or conceptualizations might be important, or what kinds of phenomena would be relevant.

As a scientific method of studying expertise, protocol analysis has by now certainly confirmed its status and proven to produce valid and reliable data (Ericsson & Simon, 1993). Two things are especially important in terms of validity. First, protocol analysis produces valid information about thought processes for concurrent verbalizations, that is, when the thinking and respective verbali-

zation occurs at the same time when the task is being executed. Second, for retrospective reports (longer than 10-30 seconds) the validity is lower and data not usable (Ericsson & Simon, 1993; Ericsson, 2006c).

Even though protocol analysis has been extensively applied, there are yet no studies concerning test-based personality assessment, which clearly is a shortcut in psychological research. With the RCS, huge number of studies has been carried out on its reliability and validity, but none on how it is actually used. Earlier it was mentioned how in medical diagnostics well working methods often do not guarantee correct diagnosis, since the assessment process brings a mental factor into play, which in turn often jeopardizes the process. This makes the research of using personality tests, even well validated ones, worth studying. This point is even more important bearing in mind that the RCS is extremely complicated to use, a shortcoming lately noticed among many eminent Rorschachers (Lerner, 1998; Meloy, Acklin, Gacono, Murray & Peterson, 1997; Meyer et al., 2001).

1.3.3 Expertise in Using the Rorschach

As noticed earlier, there are no studies of expertise and the RCS. There are however some clinician-level studies with differently experienced participants using the earlier Rorschach systems. In this section the studies that were found are reported. The first group of studies consists of studies where differently experienced participants are compared and the second group consists of studies where experienced participants are compared with some other criteria.

Studies comparing differently experienced judges. The earliest study found is that of Grant, Ives and Ranzoni (1952), who aimed to study 1) whether Rorschach experts agree on Rorschach-based judgments of adjustment, 2) whether increasing amount of information affect judgments, and 3) what kind of relationships judgments have with outside criterion if there is substantial inter-rater agreement. Additionally, participants were also to report the cues they used in ratings. There were three expert participants: A had used the Klopfer-system for ten years; B had used the Klopfer-system for six years; and C had used the Hertz-system for three years. Of the two novice participants, D was a social worker whose only experience with the Rorschach was as a subject one year earlier, and E was a graduate student in clinical psychology with one semester course in the Rorschach. The materials consisted of 146 Rorschach protocols obtained from 18-year old subjects, and participants were asked to rate these on a four-point scale (very maladjusted, moderately maladjusted, fairly well adjusted, very well adjusted). The Rorschach materials were shown to experts successively in three different phases: First only the summary page of the Klopfer and Davidson form, then this summary page plus a tabulation page showing the distribution of Variables over the ten cards, and then the earlier information plus the verbalizations (for novices only full records were shown). There were following results: 1) Experts agree with themselves and can agree with novices; 2) adding information did not change the inter-rater agreement of judgments; 3) experts' judgments did not correlate with outside criterion (inter-

view-based judgment of another social worker). Additionally, experts' cues correlated with formal signs of adjustment as well as did novice E's cues, but novice D's cues did not.

Bialick and Hamlin (1954) aimed to study how judgments of intelligence based on five Whole Responses (W) correlate with the Wechsler-Bellevue IQ. There were two groups of participants. The four more experienced participants were "staff psychologists, thoroughly trained and experienced" with the Rorschach, and who's "specific experience was both extensive and recent". The four less experienced participants were "in training". The materials were taken from the Rorschach protocols of 25 outpatients (psychiatric diagnosis of neurosis; mean age 28 years). The participants were given five arbitrarily chosen Whole Responses plus inquiry, all in one card, and they were to judge intelligence using a five point scale. All eight participants rated 25 cases. Both more and less experienced participants made valid and reliable judgments and there was no difference between groups. For the group of more experienced participants the mean correlation was .68 (SD=.11; Fisher's method of weighted z's) and reliability coefficient of two participants against two was .84. For the group of less experienced participants the mean correlation was .54 and reliability .69.

The aim of Grebstein (1963) was to 1) compare the ability of clinicians to judge Wechsler-Bellevue IQ-scores, based on structural Rorschach data, with the ability of statistical model to perform the same task, and 2) to compare the ability of three differently experienced groups of participants to perform this task. There were five participants in each group. Participants in the novice group were graduate students who had completed an introductory course on administering and scoring the Rorschach. Participants in the intermediate group were graduate students who had also completed a course on Rorschach interpretation and were in the second-year level of training. Participants in the expert group were professional clinical psychologists with a minimum of 5 years of clinical experience. All participants were trained in the Beck-system. The materials consisted of structural data of 30 Rorschach protocols that were randomly presented to each participant and they were to estimate the IQ. The statistical model was derived from 10 Rorschach factors in the Beck-system (R, W, W%, M, F%, Z, A%, P, RC (range of content)), of which the most significant factors were used to determine the best equation. The final equation was of the form $IQ = 57.54 + 3.73M + .20F+\% + .25Z + 1.20RC$. The results were 1) that there was no difference between participants and the statistical model in their accuracy, and 2) that there was no difference between the three groups of differently experienced participants. Additionally, in the most experienced group participants often used the same Variables as the statistical equation.

Turner (1966) aimed to study the interaction between the experience of participants and the amount of information needed in accurate personality description. There were four groups of participants, 25 in each, and their levels of experience in terms of the Rorschach were defined as follows: Fellows of the Society of Projective Techniques with at least 10 years of experience; graduated psychologists (PhD) with less than 5 years; graduate psychology students who

had just completed a course on Rorschach administration; and undergraduate psychology students not familiar with the Rorschach. The Klopfer-system was used. Each participant was given five protocols in each of which the information increased in the following manner: free association; plus location chart; plus inquiry and testing of the limits; plus structural data. Participants were to answer 150 true-false statements of each patient in each information level. The results indicated that 1) when level of experience increased accuracy did not improve, and 2) when level of information increased accuracy did not improve.

Studies comparing experienced judges with other criterions. Filmer-Bennett (1955) studied whether participants could correctly predict treatment outcome with the Rorschach. There were 22 pretreatment Rorschach protocols, of which half improved and half did not. Participants were 12 psychologists who "had received the Diplomate in Clinical Psychology or possessed the training and experience necessary to qualify them for the ABEPP examinations". Each of them was given 11 pairs of protocols, of which one was improved and the other unimproved. They were asked to judge which was which and give reasons for each choice. Only one participant was better than chance, and there was no consistent pattern in the given reasons.

Lisansky (1956) aimed to study inter-examiner reliability in Rorschach coding and interpretation. There were six participants who averaged 8 years of experience with the Rorschach. A ten-item questionnaire (questions on intelligence, personality traits and diagnosis) was developed to be used in the study. Each participant evaluated 20 protocols, which were accompanied by a short history abstract. Interpretation involved answering the questionnaire based on the protocols. When these were compared to control judges who made the same evaluation based on the history abstract alone, the general agreement of Rorschachers was no different from controls. However, they agreed more on sensitivity in human relationships and diagnosis, while the control judges agreed more on intellectual efficiency. In coding the agreement between participants was significantly better. There was no interaction between the reliability of interpretation and reliability of coding. In an additional task the Rorschachers listed those factors of the Rorschach they thought were relevant in answering the questionnaire items, and there was a high degree of agreement.

Armitage and Pearl (1957) aimed to examine the accuracy of Rorschach based clinical diagnosis. There were five participants with four to nine years of clinical experience with the Rorschach. The materials consisted of Rorschach structural data, verbalizations, and both of these. There were four diagnostic categories: paranoid schizophrenia, unclassified schizophrenia, neurosis, and character disorder. The participants were given 30 cases of each diagnostic class. The amount of correct judgments was not significantly higher than would occur by change. Also, none of the three groups of materials stood out as being more related to correct diagnosis.

In a study by Chambers and Hamlin (1957) the aim was to examine two questions: 1) Can participants make valid diagnosis based on the Rorschach? 2) Is there interaction between the Rorschach elements used as the basis of judg-

ments and success in the identification task? There were 20 participants all of whom had at least 3 years of experience in using the Rorschach. Some of them were very experienced, since the group included for example such eminent Rorschachers as Bruno Klopfer and Zygmunt Piotrowski. Each participant was given 5 Rorschach protocols, and the diagnostic groups comprised involutional depression, anxiety neurosis, paranoid schizophrenia, brain damage from neurosyphilis and adult mental deficiency. The task was to identify correct diagnostic group with each protocol, and give the major elements from each Rorschach that were behind the decision. There were following results: 1) Participants identified correct diagnostic classes significantly better than could occur by chance. 2) Five most successful participants used fewer words to communicate their thinking than did the six most unsuccessful ones.

Winslow and Rapersand (1964) studied the accuracy in judging whether patients improved or not after somatic therapy (electric or insulin shock), based on the contents of the Rorschach protocols. The participants were three psychologists with five or more years of experience with the Rorschach. The materials consisted of protocols from 60 inpatient schizophrenics. Participants were to use Rorschach contents in differentiating between improved and unimproved patients. Analysis showed that they were able to accurately discriminate between these two.

Albert, Fox and Kahn (1980) posed three questions in their study: 1) Can experts detect malingering of psychosis on the Rorschach by individuals inexperienced with it and uninformed about psychosis? 2) Can experts detect malingering by individuals who are inexperienced but informed by psychosis? The participants were 46 Fellows of the Society for Personality Assessment who had been Society members for an average of 20.6 years. Each participant judged four protocol sets, and they were asked to judge diagnosis, certainty of the diagnosis, dimension of pathology, and likelihood of malingering. Each 4 protocol sets included a psychotic protocol, an uninformed faker protocol, an informed faker protocol, and a normal protocol. According to the results they were unable to detect malingering. Uninformed faker protocols were diagnosed psychotic as often as psychotic protocols, and informed faker protocols were judged psychotic more often than actual psychotic protocols. Participants were equally confident in judging the four types of protocols. Fakers were also able to malinger different dimensions of psychosis. Finally, participants thought that each of the four types of protocols could equally likely be malingered.

Summary. The experts have consistently been no more successful than less experienced judges. This happened with judging adjustment or making accurate judgments based on protocols (Grant, Ives & Ranzoni, 1952), judging intelligence based on five Whole Responses (Bialick & Hamlin, 1954), judging intelligence based on protocols (Grebstein, 1963) or judging personality descriptions based on protocols (Turner, 1966). However, they were not worse than actuarial models when judging intelligence based on protocols (Grebstein, 1963). When experienced clinicians were compared with other criteria than less experienced clinicians, they sometimes judged diagnosis accurately based on

protocols (Chambers & Hamlin, 1957), but sometimes did not (Armitage & Pearl (1957); were able to correctly judge treatment outcome from contents (Winslow & Rapersand, 1964) but not from protocols (Filmer-Bennett, 1955); agreed more when judging sensitivity in human relationships and diagnosis based on protocols (Lisansky, 1956); and were not accurate in detecting malingering based on protocols (Albert, Fox & Kahn, 1980).

However, in all four of the studies that compared more and less experienced participants there is a factor that casts reasonable doubt on the nature of these studies. That is, all of them found that a participant who is not able to interpret or understand the information contents of the Rorschach can use it to make accurate judgments, which is very counter-intuitive. In the study by Grant, Ives and Ranzoni (1952) a social worker was able to judge adjustment based on the protocol; in Bialick and Hamlin's (1954) study a participant not able to interpret the Rorschach was able to judge intelligence based on five Whole Responses; in Grebstein's (1963) study a participant not able to interpret the Rorschach was able to judge intelligence based on structural data; in Turner's (1966) study participant not familiar with the Rorschach was able to use protocol to accurately evaluate personality description. These contradictions give good reasons to think that there is something problematic with these studies.

1.4 Aims of the Study and Research Questions

According to Cummings (1954), the original impetus for the above reported clinician-level studies on clinical judgment came from the problems of validity in projective methods. The idea was that even if the methods themselves were not valid enough, perhaps the validity was based on clinicians, on their skills in using the Rorschach. However, it seems that clinician-level research disappeared soon after it was born, and based on the above review it would seem natural to think that the reason for this was that they failed to show that validity of projective methods could be based on clinical judgment either. The present thesis differs from the above reported studies in three essential ways: instead of judgment the focus is on the nature of assessment process; instead of accuracy of judgment the focus is on errors in using the Rorschach materials; and instead of using less well validated Rorschach systems the well-validated RCS is used. The primary purpose of this study is to acquire new knowledge on how differently experienced clinical psychologists use the RCS in the personality assessment process, which so far has been an area lacking in empirical scientific research. Using the concepts presented in Table 1, this thesis aims to bring knowledge on one or more of the following clinician-level areas: interpretation, within-source integration, between-source integration, and personality description. While there are no studies in these areas utilizing the RCS, studies prior to the establishment of the RCS indicate that there are no differences between differently experienced clinicians, and in this sense it is to be expected that also in

this study no differences will be found. The secondary purpose of this thesis is to contrast the results to findings from the expertise research in medical diagnostics, e.g., the nature of knowledge, integratedness of knowledge, forward and backward reasoning, errors, and skill development.

2   METHOD

2.1 Design

This study follows an experimental design. Participants from three different levels of experience were given materials of two clinical cases. These materials included the anamnesis, the WAIS-R profile, and the RCS-based information. Their task was to assess the patients' personality and think aloud during the process. The method of gathering data was divided into three phases: in the first phase participants were asked to freely go through the material for about 30 minutes, in the second phase to summarize findings into a personality description, and in the third phase to explain which parts in the material their view was based upon. Respectively, these different phases were named the Free Assessment Phase, the Description Phase, and the Inquiry Phase. Originally, there was also a fourth phase, the Phase of Explaining or Understanding, but the data from this phase was not used due to problems in it. The method is an application of think-aloud method, which is based on the work of Ericsson and Simon (1993).

2.2 Participants

Initially there were 32 participants, out of whom two were excluded - one intermediate participant because of poor co-operation and one expert participant because of poor quality of recording. Of the remaining 30 participants, 21 were from Finland and nine from The United States.

All 30 participants had received special training in using the RCS. Also, through their clinical work they were all familiar with the WAIS-R. The participants were grouped based on their experience in using the RCS after their educational RCS program, as measured in years. The groups were named "expert group", "intermediate group", and "novice group". Each group included ten

participants. Nine participants in the expert group were from the United States and one from Finland. All other participants were from Finland.

Table 2 shows the information concerning participants' experience in the RCS (= the time they had used the RCS after training) and clinical experience (= the time they had done clinical work), both in years. Participants were numbered so that numbers 10-19 represent novices, 20-29 intermediates, and 30-39 experts. Novice participants had all been using the RCS for 1 year, intermediate participants between 3-9 years (M=6.2, SD=2.3), and expert participants between 9-30 years (M=20.1, SD=7.3). In the novice group the clinical experience varied between 0-22 years (M=9.5, SD=7.6), in the intermediate group between 3-34 years (M=20.1, SD=11.5) and in the expert group between 11-41 years (M=24.6, SD=8.3).

TABLE 2   Participants' Experience in the RCS and Clinical Work, in Years

	Novice group									
Participant	10	11	12	13	14	15	16	17	18	19
Experience, RCS	1	1	1	1	1	1	1	1	1	1
Experience, clinical	0	3	5	6	6	7	10	14	22	22
	Intermediate group									
Participant	20	21	22	23	24	25	26	27	28	29
Experience, RCS	3	3	5	5	5	7	7	9	9	9
Experience, clinical	7	7	3	33	34	24	30	15	21	27
	Expert group									
Participant	30	31	32	33	34	35	36	37	38	39
Experience, RCS	9	12	15	17	17	22	22	28	29	30
Experience, clinical	30	11	15	24	27	22	22	30	24	41

There were two reasons for including one Finnish participant in the expert group. On one hand, one of the American experts had to be discarded because the quality of the recording was too poor, thus leaving the expert group with nine participants only. On the other hand, the Finnish participant in question had a great deal of both scientific and teaching experience with the RCS, differing in this respect from all other Finnish participants. So, even though this participant did not differ in terms of years of practical experience with the RCS from the intermediate group, it seemed appropriate to include this person in the expert group.

All the participants were asked about the number of assessments they had done, but these seemed to be rather rough estimates and were therefore not used in grouping the participants. For example, one novice gave an estimate of 70 and one of only a few; one expert gave an estimate of 1000 assessments and

one that of 300. These numbers have some important implications. First, often the exact number seems to be very hard to define and participants end up with such rough estimates that the comparisons cannot be reliably based upon them. Second, there may be important within-group differences which arise from differences in the number of assessments done, even though the experience in years would be the same.

All participants worked as clinicians, either in hospital or outpatient settings, and/or had their private practice. In addition many of them (especially most of the experts) did personality assessments for forensic purposes, neuropsychological assessments, scientific work, and therapeutic work. All the participants did clinical assessments with either adolescents, adults, or both.

The Finnish participants were recruited by contacting the potential participants from cities nearest to Helsinki. They were told that the examiner is executing a study on how clinicians think when they assess personality with the RCS, that differently experienced clinicians were needed, that the validity of assessment was not the issue, that the study would be carried out as a think-aloud assessment task of two clinical cases, and that it would take about 90 minutes to accomplish.

Most of the American participants were recruited by email. First, the Society for Personality Assessment's (SPA) President was asked to provide a list of expert level clinicians and their email-addresses. Together with the Administrative Director they provided a list of 48 names. These were then contacted by email and given the same information as the Finnish clinicians. Seven clinicians agreed to participate by email. In addition to these, three clinicians from the same list were recruited at the SPA meeting in March, 2005, in Chicago.

With twenty Finnish participants the protocols were collected in their offices and with one at his home, between February and June, 2005. With nine American expert participants protocols were collected at the 2005 SPA Annual Meeting in Chicago, at the conference hotel rooms, and with one at a classroom after workshop.

2.3 Stimulus Materials

All participants were given the same two clinical cases to be assessed. The materials included a little bit less than one page of anamnestic information, the WAIS-R profile, and the RCS-material (Structural Summary, Sequence of Scores, Summary of Approach, Constellations Table, Interpretive Search Strategy, Responses, and Order for Reviewing Variables Within Each Cluster, stapled in this particular order). These were thought to be both familiar to all participants and roughly analogous to the everyday clinical information, excluding the use of only the profile page of the WAIS-R.

Both cases where selected from the same outpatient clinic of Jorvi, Helsinki University Central Hospital, based on following inclusion criteria: 1) The

cases had to be such that the problem/s of the patients had to be difficult to assess. The rationale was that difficult cases would bring a greater variety to results. Difficulty was defined as the inability of the staff at the psychiatric clinic to define the patient's problems. 2) Patients had to be young adults, so that both those working with adolescents and those working with adults could participate, thus guaranteeing sufficient number of participants from Finland. 3) Due to time limitations the Rorschach protocols were not supposed to be too long. 4) The information produced by the RCS had to be valid and to this end it was inspected by two examiners who determined that the protocols were correctly gathered. 5) Coding had to be reliable, and protocols were coded by two independent coders with coding reliability of 90% (percent agreement for both non-agreement and occurrence). After this the coders went through the Variables with non-agreement and made consensus decisions.

Only the profile of the WAIS-R was given, because time limits did not seem to make it possible to go through much more information, and also because the task was to assess personality, not cognitive functioning as such. However, the WAIS-R profile was included since there was no prior knowledge as to the role of neuropsychological information in personality assessment.

The first patient was a 24 year old male, and the second a 26 year old female. Both had been patients of an outpatient clinic for about two years. In what follows they will be referred to as Case A and Case B, respectively. Appendices 1 and 2 show the material that was given to participants on these patients. The information is presented here in English, noting however that the Finnish participants were given the information in Finnish. Later in the text the quotations are usually accompanied by a combination of the participant's number and the respective clinical case (e.g., "22B" or "35A").

With the American participants some problems emerged concerning the translations. In Case A, the word "homework" was erroneously intended to mean "to clean the house". In Case B it was written that the patient was the older of two children, although she was younger. In Case A there were two improper expressions, "invitation to trial" meaning "subpoena", and "distrained" meaning "debt collection" All of these errors emerged in the beginning of data collection with the American experts and were duly corrected. They did not seem to have an impact on assessments.

The materials were arranged on the table in front of the participants so that the anamnesis was on the left side (from the participant's point of view), the WAIS-R profile in the middle, and the RCS-materials on the right side.

2.4   Method of Gathering Data

Five pilot studies were done in order to develop the method to fit the research purposes. These concerned different kinds of assessment tasks (to assess personality, personality dynamics, defenses, possible psychotic features, and/or

treatment), number of clinical cases (from two to four), given time (20 or 30 minutes to freely study the materials), and finally, the convergence versus divergence of views on personality.

The different phases of the assessment task and respective instructions are shown in Figure 1. The Free Assessment Phase began with the first instruction. During this phase, and following the instruction, if the participant was quiet he/she was told to "Please think aloud!" or "What are you thinking about?" If questions were asked concerning the nature of the task, or whether participant should do this or that, the following instruction was used: "The idea is that you do it just the way you normally would". At no point were the participants allowed to use any literature, computer printouts, and so forth, since the aim was to elicit information on how they operate based solely on their knowledge structures.

Most participants were ready in about 30 minutes, usually a couple of minutes earlier or later. They then in some way informed the examiner that they were ready. They did not seem to need more than half an hour to complete the task. On three occasions about 40 minutes was used without the examiner being able to stop the process earlier.

After the participants were finished with going through and analyzing the materials, they were given the following instruction: "Could you now sum up your view on the personality of this patient". In this Description Phase the examiner wrote down all the personality characteristics included in the participant's view. On one occasion, an expert participant, having been given this instruction, wanted to go back to the materials and give it one more thought. Otherwise participants proceeded to summarize as planned. Next, in the Inquiry Phase, the examiner went through all the given personality characteristics, one by one, and asked the participant "You said that the patient is x. What in the materials do you base this on?"

In addition, a fourth phase was included in the data collection, even though in the end this phase was not used. In this phase, conducted after the Inquiry Phase, the participants were told: "How would you now, in the light of the personality assessment you just made, explain or understand the patient's problems or symptoms?" This question was intentionally multi-dimensional to see how the participants would conceptualize the material and whether they did that using hermeneutic approach, trying to understand and empathize, or in natural science approach, trying to give causal explanations, or both. However, even though this question was thought to be reasonable, it didn't seem to make sense to the participants. For almost all of the participants it was difficult to understand what they were asked to do, and this part of the data was not used.

The Free Assessment Phase
_____

> **Instruction:** *The idea is that you assess the personality of this patient, thinking aloud all the time. I hope you can do it in a maximum of 30 minutes, after which I will ask you some questions concerning the assessment you made. Here is the information on the patient: some background information, the WAIS-R profile and the Rorschach Comprehensive System materials, which consist of the Structural Summary, Sequence of Scores, Index Constellations, the Cluster Order for interpretation, Responses, and Steps for interpretation. So, assess the personality of this patient and think aloud all the time. Please start.*
> **When needed:** *Please think aloud.*
> **When needed:** *The idea is that you do just like you normally would.*

The Description Phase
_____

> **Instruction:** *Could you now sum up what is your view on the personality of this patient.*

The Inquiry Phase
_____

> **Instruction:** *Now you first said that this patient is x. What in the materials do you base this on?... Then you said that the patient is y. What do you base this on?*

The Phase of Understanding or Explaining
_____

> **Instruction:** *How would you now, in the light of the assessment you just made, understand or explain patient's problems or symptoms?*

Collecting the background information
_____

> *Thank you! Now finally I would just like to ask you a couple of questions about your education and experience.*

FIGURE 1   Structure of the Method

After the above mentioned 4 phases, the participants were given Case B, with which the procedure was identical. Two experts and one intermediate participant were not able to finish both cases during one meeting, and they were given another session.

At the end of the session the participants were interviewed about their education and experience. The interview formula is shown in Appendix 3. In most cases it was done immediately, but due to time restrictions some subjects wanted to answer it by email. Finally, all the participants were given a written certified statement on confidentiality, where the examiner guarantees that participants' identities will not be revealed in any way. This is shown in Appendix 4.

A total of 64 protocols were collected (32 participants x 2 cases), and as noted previously, protocols from two participants were not included. In the end, there were 60 protocols from 30 participants, divided into three experience groups. These protocols comprised totally 610 pages, of which Case A included 325 pages and Case B 285 pages. In terms of the four different phases, the Free Assessment Phase consisted of 345 pages, the Description Phase 56 pages, the Inquiry Phase 168 pages and the Phase of Understanding or Explaining 41 pages.

There still remained some incompleteness in the data, all because of problems in recording. Regarding protocols 10A and 10B, only the Description Phases could be utilized, and protocol 16B does not exist because of a problem with the tape recorder. All these shortcomings were taken into account in the statistical analyses. Thus, although there were totally 30 protocols, when these problems were noticed, in the statistical analyses of the Free Assessment Phase there were 28 protocols, of the Description Phase 29 protocols, and of the Inquiry Phase 28 protocols. Regarding protocol 36A, between 2-5 minutes of the Free Assessment Phase are missing, but this protocol was included as such.

2.5 Categories

This study represents a data-driven approach to analyzing the process of RCS-based clinical personality assessment. Data-driven approach, often also referred to as inductive or bottom-up approach, is a widely used approach when the aim of research is to generate hypotheses in a new research area where there does yet not exist a well-developed theory (e.g., Breakwell, Hammond & Fife-Schaw, 2000; Cavanagh, 1997; Dey, 1993). On a general level, in this kind of approach the data is first reduced by excluding irrelevant information, then clustered into categories, and finally abstracted into more theoretical units (Cavanagh, 1997; Dey, 1993). In the initial stages of starting to create the categories, it is advisable to take a sample of the data (e.g., Richardson, 1996). This procedure was applied with the categories where the identification was not reliable.

2.5.1 Items

An item is any piece of information that was included in the given materials - in the anamnesis, the WAIS-R, or the RCS. The analysis of items brings understanding about how the materials were used. The use of the materials was estimated on the basis of the Free Assessment Phase, where the assessment process started by the participants scrutinizing the given materials, and on the basis of the Inquiry Phase, where the participants gave grounds for the personality characteristics they had given. Items from these two phases are defined a slightly differently.

Free Assessment Phase. An item from the RCS was e.g. a Variable (e.g., *Human Movement*), Index (e.g., *Coping Deficient Index*), a Variable constellation (e.g., *M:W, FC:CF+C*), Response, or a combination of Variables (e.g., *FQ-, DdS+99*, or *Blends*). In terms of the Responses, reading all of a specific Response or just some part of it would both be regarded as one item. With regard to the WAIS-R profile, an item would comprise e.g. a specific scale (e.g., *Digit Symbol*), a cluster of scales (e.g., *Performance Scales*), scatter, profile, or a numerical value from the WAIS-R (e.g., *112*). If the participant at the same time verbalized a scale and the value, only one item was included (e.g., *Verbal Scale is 112*). Gen-

erally, if the same item was mentioned more than once only one occurrence was included.

In the following excerpt from participant 33B, in the beginning participant goes through some of the WAIS-R, and later through some of the RCS-materials:

> *er **99 102 101** er so er you know average intelligence er we are looking at a **verbal abilities** they're all pretty uniform with the exception of **similarities** … got **17 Responses**, that's that's you know find little low but you know it's okay, er **Lambda is at .13** that's low, er suggests somebody who's maybe a little bit over whelmed, you know you always want a little bit more filtering and so er that may be problematic, er look at the **suicide constellation** and and I see some problems here obviously*

In the episode concerning the WAIS-R, participant 33B mentions items 99, 102, 101, verbal abilities (=scales), and similarities. Concerning the RCS-materials three items are mentioned: 17 Responses, Lambda is .13, and Suicide Constellation.

Items from the anamneses during the Free Assessment Phase were not included in the analyses for two reasons. First, since all participants read aloud all of the anamnestic information, the number of items was the same. Second, it was not possible to define clearly what exactly would comprise an anamnestic item. For example, would it be a word or a sentence? Hence, only items in the RCS and the WAIS-R profile were counted in.

Inquiry Phase. This Phase consisted of the examiner asking grounds for the given personality characteristics. The following dialogue between the examiner and participant 35B is an example of one inquiry episode. Here the examiner inquires what in the materials made the participant think that the patient has no good self-esteem:

> *E: okay, how about **no good self esteem**?*
>
> *P: okay, the **history** would suggest that, er and er the E... er **Egocentricity index** would suggest that er and then the er **PHR** being so high would suggest that, er and er that's my main points*

Participant 35A responds, that the idea of low self-esteem is based on the history, Egocentricity Index, and the PHR. In the following example novice participant 14B gives grounds for personality characteristic aggressive:

> *there are very many **S-Responses**, eight … **GHR:PHR 11** … er in special scores **three aggressions***

Thus, the Variables Space, GHR:PHR and AG make this participant think that the patient is aggressive.

When participants verbalized the anamnesis, sometimes they mentioned generally the anamnesis, and sometimes a specific part of it. In both cases one item was included. Also, if different parts of the anamnesis were cited to justify same conclusion, only one occurrence was included. With regard to the RCS, an item was defined similarly as concerning the Free Assessment Phase. In terms

of the WAIS-R, only one occurrence was registered even when different parts of it were verbalized.

2.5.2 Inferences

An inference is here defined as a mental content that the participant brings to the personality assessment process, and which is about the patient – not materials or the participant himself. An inference was identified when a participant verbalized something that was not written in the materials, and thus had to be based on mental contents. Inferences were identified in the Free Assessment Phase only, since only then inferences in this sense appeared. During the Description Phase participants listed the personality characteristics, and during the Inquiry Phase they noted the relevant items.

The category of an inference has been stipulated in many studies concerning medical expertise (e.g., Patel & Groen, 1991a). The basic idea is that the data can be divided into two parts: that which repeats the given materials and that which adds to them based on the mental contents of participants. Patel and Groen (1991a) define it as a transformation made on the proposition in the materials. This non-given element can theoretically be named in many different ways, e.g., as an inference, hypothesis, non-given, etc., none of which is totally satisfying. The concept of an inference has been chosen here as perhaps the best alternative, but it nonetheless brings a possibility for misunderstanding. Most importantly, while it indicates elements in the data in which the participant has formed a mental content based on the given materials, it still is not fully equivalent to the concept of inference in logic or in everyday language.

Since there were three different information sources, they were all used as a basis for making inferences. Therefore, there were inferences related to anamnesis, inferences related to the WAIS-R, and inferences related to the RCS. It was usually easy to see where in the materials they were derived from, since the verbalizations mostly proceeded from materials to inferences. The following episode from participant 33B, already analyzed above in terms of items, is now analyzed in terms of inferences attached to the WAIS-R and the RCS:

> *er 99 102 101 er so er you know* **average intelligence** *er we are looking at a verbal abilities* **they're all pretty uniform with the exception of** *similarities … got 17 Responses, that's that's you know* **find little low but you know it's okay**, *er Lambda is at .13* **that's low**, *er* **suggests somebody who's maybe a little bit overwhelmed**, *you know* **you always want a little bit more filtering and so er that may be problematic**, *er look at the suicide constellation and and* **I see some problemshere obviously**

In this episode, using the WAIS-R profile materials, participant forms an inference based on *99 102 101*, another one from *verbal abilities* and *similarities*; then from the RCS-materials makes inferences from R=17, L=.13 and S-CON.

There were episodes where it was not possible to see reliably the respective item or source that functioned as a starting point for an inference. For these cases a theoretical entity called "unspecified source" was hypothesized. There may have been a specific item in the mind of the participant, but it was not ex-

plicitly verbalized. The following example is from participant 34B, where the first inferences are formed from the RCS, but the latter ones are different:

> *W to M 10 to 12 so she's **not able to neutralize all her strengths very effectively**, er well of course she's depressed so she wouldn't, 3 AG 0 COPs, er, **not very easy to get along with in the world**, I mean she left her job cause she had trouble with another worker*

In the beginning, Variable Constellation W:M is a basis of forming two inferences. Then it seems that AG and COP are the basis for an inference *not very easy to get along with in the world*, but the words after that indicate that it is also formed based on those anamnestic pieces of information, and thus this belongs to the class "unspecified source".

Since the problem-solving task was to assess personality, most of the inferences were specifically about what the patient is like. However, participants often paid attention also to the patients' environmental factors as a part of personality assessment, and hence these episodes were included in inferences.

There appeared an important technical question as to how to identify a single inference, that is, when a sequence of reasoning was one inference and when more than one. The following episode from 36B illustrates the point:

> *part of the pain of this lady is that **she has sort of a normal desire for er affection and closeness** but er **other people are so scary for her** and so **completely mysterious to her**.*

The basic component of an inference is an idea about what the patient is like. This inference as such is too complex to form one inference. When it is analyzed into more basic components, it has three inferences: 1. *part of the pain of this lady is that she has a normal desire for affection and closeness.* 2. *but other people are so scary for her* 3. *and other people are completely mysterious to her*. If the verbalization had instead been

> *part of the pain of this lady is that **she has sort of normal desires for affection** and **aggression towards other people**, but er **other people are so scary for her** and so **completely mysterious to her**,*

it would have had four inferences. Verbalization *desires for affection and aggression towards other people* has two very different ideas, affection and aggression, and has hence two inferences. The earlier verbalization *desire for er affection and closeness* however was treated as one inference.

Further, defining the identity of an inference was important since the participants often produced similar verbalizations in two different episodes. The following technical definition was stipulated. If the participant repeated an inference twice, then these were treated as one inference. If the inference appeared in two different episodes with some other episodes between them, they were treated as two inferences. The rationale behind these definitions was twofold. First, it was possible that the episodes between the two inferences had an impact on the meaning of the verbalizations. Second, if the two inferences did not have identical premises, they would not be identical.

There existed a sub-class of inferences, called predictions. While participants almost always started the assessment process by reading the anamnesis, it was easily noticed that during this some participants started to make hypotheses as to how the respective RCS-materials would look like. These episodes were accounted in the analyses. The first example is from 36A who infers from the history how test results may look like:

> *fact that he tells things in an organized way suggest that probably there isn't thought disorder er no impression of any psychotic features.*

In the next example participant 39A tells what the RCS should look like in terms of coping, because the patient has problems in managing everyday life:

> *but he can't work now so he's regressed really he is much more like a preteen or even younger who who needs to have his mother take care of him, he er he is of course depressed er and he can't cope he has a coping deficient problem and we should see on the R we should see the CDI er up.*

2.5.3 Integration

Integration means that two pieces of information are connected. The combined information may consist of items or inferences. Therefore integration is a phenomenon that conceptually is on a different level than items and inferences. There were two types of integration that are defined differently: integration between information sources and integration within a specific information source (the RCS).

Between-source integration. In the Free Assessment, the first type of integration is that of a combination between sources of information: between the RCS and anamnesis or the WAIS-R, or with all three at the same time. Participants went through the information sources (anamnesis, the WAIS-R profile, and the RCS-materials) one at a time, and hence between-source integration episodes were identified when the participant momentarily started to talk about another information source and combined the two. A mere move from one information source to another is not integration, as integration requires some verbalization to signify the activity of integration.

In the next episode, 32B integrates anamnesis with D and AdjD:

> *look **D of +3 AdjD of +4** er the s… I can't believe that given this **history** er she falls apart every time she gets involved with a woman and it immobilizes her*

32B reasons from the anamnesis that the patient *"falls apart every time she gets involved with a woman and it immobilizes her"*. In other words this is an inference related to an item in the anamnesis. This is thought by 32B to be in contradiction with D=+3 and AdjD=+4. The activity of integration is signified by the verbalization *"I can't believe that given this"*.

Another example of integrating D and AdjD with the anamnesis is from participant 33B, who integrates inferences from these Variables with materials from anamnesis:

*but with this **D of +3** and this **AdjD of +4** we talk about somebody I mean it's very interesting because it's it's on one level sort of things contradictory **here is somebody who is saying I'm suicidal and I'm overwhelmed** and yet you got a a D and AdjD that that would suggest that they are er you know er **very resistant to stress.***

Also according to 33B these information sources are in contradiction. 33B repeats the material indicating idea that the patient is suicidal and overwhelmed, and related to D and AdjD forms the inference *"very resistant to stress"*. These are combined with the verbalization *"it's on one level sort of things contradictory"*.

In the Inquiry Phase between-source integration was defined a little bit differently. There participants provided the items in the materials on which the personality characteristics were grounded, and between-source integration was identified when a ground included an item from the RCS and anamnesis or the WAIS-R. In the following example the examiner gives the personality characteristic and asks participant 35B for the grounds, after which the participant gives materials from all three sources:

E: okay then you mentioned that she hasn't er er good relationships, where do you base that?
*P: **history**, er okay, er I forgot to mention the **similarity score**, er if you're not able to abstract what you're experiencing er you're probably thinking about things in a concrete way and so the way you're imposing your ideas on other people and the feedback that you're getting might be confusing to you er the relationships **no Cooperative**, 3 **Aggressive**, 7 **Space** minus, **Egocentricity**, er the **H:(H)** Responses er and then I I'm sure I mentioned some stuff in the **content** related to that as well but I would stay with primarily in the **Structural data**.*

Thus, behind characteristic *hasn't good relationships* there were the anamnesis, one item from the WAIS-R, and five items from the RCS.

Within-source integration. In the Inquiry Phase it was possible to identify grounds that included within-source integration of materials from the RCS. These were grounds where the participant mentioned more than one RCS-item. Thus, the same episode may include both between-source and within-source integration. For example, the earlier example from 35B has materials from all three sources, but since it has five items from the RCS, it thus at the same time includes within-source integration.

2.5.4 Errors

Errors give information about how materials were understood. They were identified with regard to the inferences, integration, and use of the Variables in grounds.

Errors with inferences. In the Free Assessment Phase, sometimes the verbalization of an inference related to a Variable made it evident how the participant understood that specific Variable. It is important to notice here that, based on the definition of an inference made earlier, inference does not necessarily mean something that is inferred from materials, but only a mental content that it verbalized related to some material.

The errors in inferences were first identified by two examiners, who jointly examined one protocol from each group. After the phenomenon was found, these examiners jointly examined all protocols from Case A, and identi-

fied all errors. In these cases, to evaluate the correctness of the use of the Variable the criteria were taken from Exner's book, A Primer for Rorschach Interpretation (2000). Then two examiners independently evaluated 12 randomly chosen protocols from Case B (3 from each group), and the kappa coefficient was .35, indicating fair agreement (Landis & Koch, 1977). Then the rest of the protocols were evaluated, and findings were statistically analyzed.

Based on their contents, the errors were further analyzed into two classes: numerical and conceptual errors. Also this division was jointly identified by two examiners. A numerical error means that the participant verbalized an incorrect quantitative interpretation of a Variable. In Case A EA=6.5, which indicates normal amount of resources, but in the following example participant 24A erroneously thinks that the amount of resources is low:

> *when you look at the EA you can see that right now there are **not very many** re sources*

A conceptual error in turn means an incorrect understanding of what a Variable means. In the following example participant 22B incorrectly understands M-=6:

> *her M- is 6, there can be **escaping to fantasy**, she **substitutes reality with fantasies with imagination and this way avoids confronting problems***

The logic of 22B is very understandable in the sense that M- signifies ideation that is unrealistic, which in turn could be understood as meaning that as a way of coping the patient substitutes outer reality with inner fantasies. However, based on Exner (2000) M- means strange thinking, and using fantasy as a defense could be reasoned from the Variable Constellation Ma:Mp, not M-.

Errors in between-source integration. In the Free Assessment Phase, the between-source integration episodes were divided into two classes, correct and incorrect. This phenomenon was identified by two examiners. After jointly examining episodes and noticing that there were different reasons for the integration to collapse, a decision tree was created (Figure 2) to rate these episodes accordingly. It starts from the used Variables, from whether they are used correctly in the episode. If not, then the episode is incorrect – otherwise the evaluation proceeds to examine the second part, and whether there the anamnesis or the WAIS-R is interpreted correctly or not. Finally, the connection between these must also be interpreted.

Has the test Variable been interpreted correctly?
→ No → Integration is incorrect.
→ Yes → Move to next phase.

Has the anamnesis (or the WAIS-R) been interpreted correctly?
→ No → Integration is incorrect.
→ Yes → Move to next phase.

Is the RCS correctly connected to the anamnesis (or the WAIS-R)?
→ No → Integration is incorrect.
→ Yes → Integration is correct.
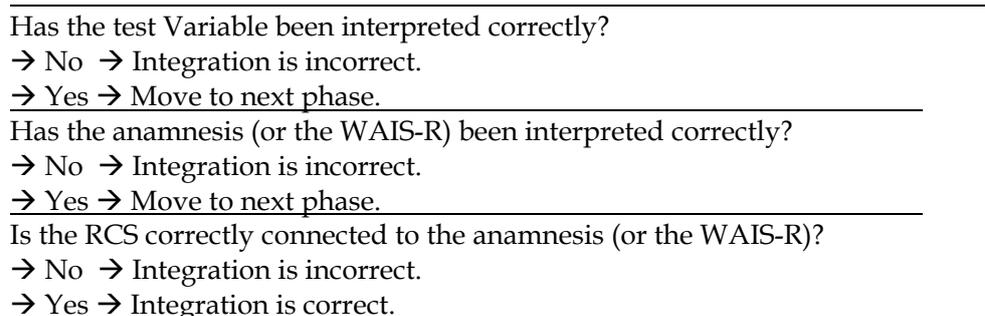
FIGURE 2   Decision Tree for Assessing the Correctness of Between-Source Integration

Theoretically, evaluating the correctness of anamnesis is more difficult than other sources in the sense that there are no norms. But practically that was not the case since it was easy to say when anamnesis was rephrased in some strange way. In the following example participant 11A integrates Indexes with the anamnesis:

> none of the Indexes is positive which is **not a surprise in light of the anamnesis**

In this episode there are no inferences, and integration is about combining the RCS and anamnesis. The integration in this example is incorrect, since there are at least two Indexes that one could expect to be positive based on the anamnesis: The patient has problems with other people in everyday life, and it was mentioned that he is depressed. One could thus suppose that CDI and DEPI were positive.

In another example from participant 22A again Variables are connected to anamnesis, but now the problem is that anamnesis has been incorrectly interpreted:

> well a is higher than p that's a good thing, he is not like escaping to fantasy in his thinking, or Ma is higher than Mp, you cannot think he would do that, and there are no signs of passive dependency, **even though he has regressed to live in his mother's place**

The inference formed from the RCS is correct, although the participant seems to hesitate with the interpretation. But the way anamnesis is interpreted is incorrect: Anamnesis says that the patient gave up his apartment to pay his parents' debts, and therefore had to move to his mother's place. But participant interprets this to mean that the patient is regressed to live at his mother's place, which is not true.

With the decision-tree, two examiners independently rated the episodes of 12 randomly chosen protocols from Case B (4 from each group), with the kappa coefficient of .82, indicating almost perfect agreement (Landis & Koch, 1977). Then the episodes from other protocols were rated and results analyzed statistically.

Errors in grounds. In terms of errors in grounds two different categories were identified jointly by two examiners while examining the protocols from Case A. First, sometimes errors were about using the RCS-variables to argue for a personality characteristic, even though it could not be used to ground that kind of a personality characteristic. This kind of error was labeled "inappropriate use of a Variable". In the following example participant 14A gives grounds for personality characteristic *signs of being inhibited*:

> well it is the **anamnesis** and er well from the Rorschach there are indications like for example this **GHR:PHR COP** is 0 **AG** is 0 like indications that he does not experience social activity being very satisfying

The problem here is that *signs of being inhibited* cannot be evaluated with the RCS, and hence it is here used inappropriately. Two examiners independently

evaluated 12 randomly selected protocols (4 from each group, including 129 items), and the kappa coefficient was .41, indicating moderate agreement (Landis & Koch, 1977).

Second, sometimes an error was about using a Variable incorrectly. This means that it was possible to use the RCS in grounding some personality characteristic, but when that was done an incorrect Variable was used. An example of this incorrect use of the RCS would be an example from 29A in grounding for a characteristic *alienated*:

> *that was in perception er there is **DQ vague X unusual** er **M minus***

Alienated is a characteristic than can be estimated with the RCS, but it is specifically linked to Space Responses. DQ vague and M- are not as such indications of alienation, but Xu% would be closer to the idea of alienation when it is high. Two examiners independently evaluated 12 randomly selected protocols (4 from each group, including 132 items), where the kappa coefficient was .53, indicating moderate agreement (Landis & Koch, 1977).

2.5.5 Content of Personality Characteristics

In the beginning of the Description Phase the examiner asked the participants to give a description of the patient's personality. The whole response to this instruction is called a personality description, and the individual items that a personality description is composed of are called personality characteristics. Personality descriptions were broken down into personality characteristics by two examiners following a consensus-principle, after which they jointly clustered them into categories.

The first categories that were formed in analyzing personality characteristics were those about their content. Content stands for the primary matter that a characteristic is about. As an example, here is the whole personality description of 37A which is broken down into personality characteristics (in bold and numbered):

> *this is **1) a young man whose history er has er predisposed him to difficulties**, who **2) has done better when in structured er setting**, **3) when the structured setting has been taken away from him through family divorce he tries to fill the role of providing for mother who is the victim** er **4) but he is inadequate to do that, based on his inability to stay with his education 5) and probably making some bad judgments about his er lifestyle**, er so th… **6) so there is some judgment problems, 7) his thinking is er somewhat er immature and er rather unsophisticated, 8) he has chosen to deal with the problems in his life by withdrawal and isolation**, … **9) there's is no clear depression, 10) there's no clear thinking disturbance**, er but **11) there is this primitive quality to his makeup that is likely to er undermine him in the future**, I'm not sure how much more to say, yeah I mean I think I have said about what I would like to see in terms of treatment, **12) what I would like to see in terms of treatment is structure er in his life**, er I always ask myself do I like this person, do I not like this person, er I always ask myself do I want to work with this person, is this somebody that I'd think I would like to work with because I like his personality, I like something sort of there, or is this something that I think would be difficult to work with, er **13) should not be an easy person to work with this is going to be longer term er because of of having some primitive personality makeup**, I also am concerned er about the family history and about the drinking that he is doing and that*

will make things more difficult, er and I guess from my perspective er he would be okay to work with, he would be er I don't have to put a lot of energy into working with him, so er I'm not sure if I wanna work with him or not, I'm not sure how much I like him, he's okay I guess but I don't know that I'd wanna work with him right now

This personality description from 37A consists of 13 personality characteristics. As can be noticed, there can also be other things verbalized during this phase that do not count as personality characteristics.

The personality characteristics included multiple diverse contents, which were analyzed. Some themes appeared repeatedly, and these formed the basic structure of starting to classify them. The classification system underwent many revisions and at the same time explicit instructions for coding were developed (see Appendix 6). Table 3 shows the classification system that was finally obtained. The codes are shown in brackets. Personality characteristics were divided into five broad classes: Internal, External, Diagnostic, Treatment, and Other personality characteristics. Then the first three were divided into subclasses. Internal characteristics were divided into following subclasses: Experience (what the patient experiences), Cognitive (e.g. thinking, reality testing, ways of perceiving, information processing), Neuropsychology (e.g. intelligence, abstract reasoning, noticing visual cues), Coping (e.g. problem-solving, defenses, resources, controlling, regulating), Relation to Oneself (different kinds of ways to relate to oneself), Motivation (motivations to internal or external action), and General characteristic of personality. External characteristics were divided into Behavior (descriptions of patient's behavior), Relationship (descriptions of relations to other people), and descriptions of External Circumstances. Diagnostic characteristics included descriptions of psychiatric diagnosis or phase of diagnosis, and Treatment characteristics evaluations of treatment. A residue class Other characteristics was established for those that did not fit any other class. Altogether, the system consists of 14 different classes.

TABLE 3   The Classification System of Personality Characteristics

Main Class	Sub-class
Internal	Experience (EXP)
	Coping (COPE)
	Cognitive (COG)
	Neuropsychology (NP)
	Relation to Oneself (SELF)
	Motivation (MOT)
	General (GEN)
External	Behavior (BEH)
	Relationship (REL)
	External Circumstances (EC)
Diagnostic	Psychiatric Diagnosis (DG)
	Phase Diagnosis (PDG)
Treatment	Evaluation of treatment (TREAT)
Other	Other (O)

Initially, two examiners independently coded three protocols from Case A, but their agreement was very poor. Therefore, explicit coding instructions were developed (Appendix 6), and using these, two examiners independently coded 141 personality characteristics (4 randomly chosen protocols from each group in Case B). The kappa coefficient of coding was .74, indicating substantial agreement (Landis & Koch, 1977). Then the rest of the summaries were coded and findings were statistically analyzed.

The personality characteristics from protocol 37A's Description Phase are coded here for illustrative purpose. Personality characteristics stem from the personality description, and codes are in brackets.

- *this is a young man whose history has predisposed him to difficulties (EC)*
- *who has done better when in structured setting (COPE)*
- *when the structured setting is taken away from him through family divorce he tries to fill the role of providing for mother who is the victim (REL)*
- *but he is inadequate to do that based on his inability to stay with his education (COPE)*
- *and probably making some bad judgments about his er lifestyle (COG)*
- *so there is some judgment problems (COG)*
- *his thinking is somewhat immature and rather unsophisticated (COG)*
- *he has chosen to deal with the problems in his life by withdrawal and isolation (COPE)*
- *there's is no clear depression (EXP)*
- *there's no clear thinking disturbance (COG)*
- *but there is this primitive quality to his makeup that is likely to undermine him in the future (GEN)*
- *in terms of treatment I would like to see structure er in his life (TREAT)*
- *should not be an easy person to work with, this is going to be longer term er because of of having some primitive personality makeup (TREAT).*

This personality description includes 13 personality characteristics from three different main classes (Internal, External, Treatment), and from seven different subclasses (COPE, COG, EXP, GEN, EC, REL, TREAT). The participant mostly characterizes patient A's personality in terms of Cognitive and Coping characteristics.

2.5.6 Structure of Personality Characteristics

While analyzing the contents of personality characteristics it was noticed that characteristics with similar contents may have been verbalized differently. For example, the following instances are all about depression (content class: EXP), but there are clear structural differences:

depressed (14A, 15A, 23A, 26A, 27A, 28A, 32A), *reactive depression* (24A), *the use of alcohol causes depression to increase* (25A), *chronic depression is associated with pervasive sense of anger* (31A), *trying to get out of schizoid situation which generates anxiety and depression* (31A), *not reactive depression* (33A), *is a good son by being depressed and anxious*

(34A), *a major shutdown caused depression* (35A), *prone to depressive episodes* (35A), *no clear depression* (37A), *evidence of depression* (38A).

Seven of these participants used the same form, *depressed*, but all the other formulations were structurally different. They are different in terms of length, but also in terms of how depression is described: they may describe the nature of depression, the cause of it, its function, or tendency towards it.

The structural analysis was executed using concepts from formal logic. The idea of using formal logic to analyze natural language is not new in science. In philosophy one of the most known examples is from Russell (1905/1956), who in his theory of definite descriptions analyzed the logical structure of sentences to better understand their semantics. The reason for this came from a paradoxical situation in philosophy of language, namely, how is it possible that a sentence about some individual can make sense when the actual individual it tells about does not exist. In philosophical analyses the example was the sentence "the present king of France is bald". To solve this paradox Russell used formal logic to analyze these kinds of sentences. After analyses, the above example turned out to have the logical form of the sentence: "there exists an individual, and this individual is the present king of France and this individual is bald". Thus, the original sentence is not about an individual, but about the existence of an individual. There is no paradox, since the sentence does not claim that a non-existent individual has some property, but claims that there exists an individual and that individual has that property. This sentence of course is not true. Russell's analysis contributed fundamentally to the 20th century philosophy of language and the so called analytic philosophy. Analyses of the surface structure of natural language to reveal its deep structure have appeared also in other areas of science. For example, in his transformational grammar, Chomsky (1957/1985) analyzed the grammatical surface structure of natural language sentences and ended up with the universal grammar.

This thesis uses concepts derived from first-order predicate logic to analyze the structure of verbalizations of personality characteristics. Predicate logic is a formal language the purpose of which has been to enable talking about individuals and their properties in formal logic. A standard language of first-order predicate logic (e.g., Suppes, 1957) includes individual variables (x, y, z, …), predicates that are properties of individuals (e.g., P), conjunction & ("and"), disjunction V ("or"), implication → ("if … then"), equivalence ↔ ("is equivalent with"), = ("is identical with"), and ¬ (negation). In addition, there are left and right parentheses. There are also two quantifiers ∃ ("there exists") and ∀ ("of all x it is true that"). With this calculus, e.g., a sentence "Plato is Mortal" could be transformed into "∃x(Px & Rx)", meaning, "there is someone, who is Plato and who is mortal".

When initial transformations of verbalizations with this formal language were applied it was noticed that some modifications had to be made for it to fit the purpose of analyzing the structure personality characteristics. 1. Quantifiers were both useless and difficult to use, and they were excluded. 2. Participants sometimes verbalized contexts from the patient's life (example: *this difficult*

situation makes him drink), and a concept "c" was adopted to signify these instances, since the aim was not to analyze more thoroughly contents repeated from the materials. 3. Verbalizations were sometimes about causal relationships (example: *drinking has caused anxiety and depression to increase*), and a concept (→) was adopted for these while at the same time the implication of formal logic was excluded. 4. The variables of formal logic (individual variables and predicates) were substituted by two constants, an individual term (i) and a predicate term (P). 5. The predicate term was defined as indicating both an adjective (*anxious*) and a verb (*tries to, started to drink*). 6. An additional concept modifier (M) was included to signify the idea of to what extent something exists. 7. There seemed to be no use for the disjunction symbol V which was excluded. Eventually, the following eight elements were used:

- Term (i) (examples: *depression, thinking, thought disorder*)
- Predicate (P) (examples: *depressed, disordered; think, repeat*)
- Modifier (M) (examples: *very, little, much*)
- Negation (~) (examples: *no, not*)
- & = conjunction (*and*)
- → = causal relationship (example: *the abuse of alcohol causes depression*)
- c = context (example: *these circumstances have exposed him to depression*)
- left and right parenthesis.

Using these elements the structures of internal personality characteristics were analyzed. Illustrating the structural analysis, the internal personality characteristics from 37A are coded below. In the examples, the first sentence is the original one. The second sentence explicates the structure of the first one, and the third sentence translates those structures using formal language.

- *has done better when in structured setting*
  = *(when in structured setting) causes (better (does))*
  = c → MP
- *he is inadequate to do that based on his inability to stay with his education*
  = *(not (able to stay with education)) = (not (is adequate (to do that)))*
  = ~c → (~PP)
- *and probably making some bad judgments about his er lifestyle*
  = *(making (bad (judgments))) lifestyle*
  = PMPi
- *so there is some judgment problems*
  = *(problems in (judgment))*
  = Pi
- *his thinking is somewhat immature and rather unsophisticated*
  = *(somewhat (immature (thinking))) and (rather (unsophisticated (thinking)))*
  = MPi & MPi
- *he has chosen to deal with the problems in his life by withdrawal and isolation*
  = *(problems) causes ((withdrawal) and (isolation))*

$$= P \rightarrow (c \; \& \; c)$$

- *there is no clear depression*
    = (not (clear (depression)))
    $= \sim Mi$
- *there's no clear thinking disturbance*
    = (not (clear (thinking disturbance)))
    $= \sim Mi$
- *there is this primitive quality to his makeup that is likely to undermine him in the future*
    = (primitive (quality (makeup))) causes (undermine (in the future))
    $= PPi \rightarrow Pi$

These nine personality characteristics from participant 37A include some simple structures and some more complicated. There are eight terms (i), twelve predicates (P), six modifiers (M), four implications ($\rightarrow$), four contexts (c), two conjunctions (&), and four negations (~). There were also several parentheses but they are not included in the analyses.

2.6   Statistical Analyses

The data was first analyzed qualitatively in the sense that the above categories were formed. The findings were then analyzed statistically. Cases A and B were analyzed together. For the normally distributed variables, the repeated measures of ANOVA was used to compare mean differences between groups. If the group effect was significant, the pairwise comparisons were made using Bonferron test. Otherwise, for the non-normally distributed variables, the nonparametric Kruskal-Wallis test with exact p-value was calculated (if the time limitation of five minutes for exact test was not enough the Monte Carlo p-value was calculated). If the group effect was significant, the pairwise comparisons were made using nonparametric Mann-Whitney U test.

3   RESULTS

In this study the problem-solving process of the RCS-based personality assessment consists of three phases: In the Free Assessment Phase participants freely go through the materials and evaluate them; in the Description Phase they describe patient's personality; and the Inquiry Phase they give grounds for the personality characteristics they had given. This chapter presents the results from these phases in this order.

3.1   The Free Assessment Phase

In the Free Assessment Phase participants were given about 30 minutes to study the materials. During this time the main instruction was to assess personality and think aloud. They were allowed to work as they wanted, except that they had to think aloud all the time and base the assessment solely on the given materials. The presentation of results from this phase divides the analyzed phenomena according to the categories presented in Chapter 2: items, inferences, integration, and errors.

3.1.1 Items

Since participants started the assessment by going through the materials (anamnesis, the WAIS-R profile, and the RCS-materials), it is natural to begin by examining how they used items that these sources included. Materials from the WAIS-R and RCS are analyzed, but the anamnesis-based items are not since participants read all of the anamneses, and therefore the number of anamnestic items was the same in all groups. The WAIS-R items were used in total 323 times and the RCS items 2325 times. The way these were used by differently experienced clinicians is presented in Table 4. Although all items were used more by experts, there were no statistically significant differences. Items from the WAIS-R profile and from the RCS (both Variables and Responses) were used similarly.

TABLE 4   The Number of Items from Two Sources, the RCS and the WAIS-R

Source of Item	N		I		E		F	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
WAIS-R	5.3	3.9	4.9	4.0	7.0	4.8	1.39[a]	2,25	.268	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
RCS-item	36.3	14.6	38.7	16.9	48.6	14.1	1.91[a]	2,25	.169	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Variable	26.9	14.2	28.5	11.9	38.2	14.1	2.08[a]	2,25	.147	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Response	9.4	5.9	10.2	7.1	10.4	6.8	0.08[a]	2,25	.927	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test.
Statistical significance: **p<.01, *p<.05.

These statistics also show that participants did not go through all of the materials. The number of Variables in the RCS is much higher than the above means, and also the number of Responses in Cases A and B was much higher than the means in Table 4 (19 Responses in Case A and 17 Responses in Case B). The lower number of WAIS-R items, compared to the Variables, most probably resulted from the fact that the participants were given only the profile page of the WAIS-R. However, even if all of the WAIS-R materials were given, the means might have been lower since the task was to assess personality, not cognitive functioning.

3.1.2 Inferences

Items in all three sources were used to make inferences in terms of what the patients were like, or what their environment or history was like. As mentioned earlier, items obtained from the WAIS-R were used in total 323 times. From these 283 inferences were made. Similarly, while the RCS-items were used 2325 times, from these 2565 inferences were made. Thus, an average of 0.9 inferences was formed related to a single WAIS-R –item, and 1.1 to an RCS-item. Furthermore, there was an additional source "unspecified", since it was not always possible to recognize what the inference was connected to.

The number of formed inferences was analyzed statistically, and the results are shown in Table 5. The number of inferences made from the anamnesis was similar in different groups. Compared to novices, experts made significantly more inferences from the WAIS-R, the RCS-variables, and Responses. In addition, experts made more inferences from the Responses than intermediates. With unspecified source no differences were found. Further, there were no differences between intermediates and novices in any respect.

TABLE 5   The Number of Inferences from Items in Different Sources

Source	N M	N SD	I M	I SD	E M	E SD	F/Chi	df	Sig	Pw	Sig
Anamnesis	8.3	8.5	9.3	9.5	14.0	12.2	1.81[a]	2	.416	N<I[b] N<E[b] I<E[b]	ns ns ns
WAIS-R	3.2	2.3	4.4	6.5	7.2	7.3	5.65[a]	2	.054	N<I[b] N<E[b] I<E[b]	ns .028* ns
RCS-item	27.3	20.5	44.4	22.4	62.1	37.4	6.84[a]	2	0.25	N<I N<E I<E	ns .021* ns
Variable	25.0	19.6	41.6	21.5	53.1	33.6	3.41[c]	2,25	.049	N>I[d] N<E[d] I>E[d]	ns .045* ns
Response	2.3	2.7	2.8	3.0	9.0	8.4	6.93[a]	2	.027	N<I[b] N<E[b] I<E[b]	ns .022* .030*
Unspecified	16.7	21.3	16.2	12.7	16.1	14.5	1.31[a]	2	.862	N>I[b] N>E[b] I>E[b]	ns ns ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test. [c]F test. [d]Bonferroni test.
Statistical significance: **p<.01, *p<.05.

Thus, making inferences is an activity that characterizes experts' assessment process. While the number of items was the same with both Variables and Responses (Table 4), experts' higher number of inferences from these materials indicates that they are utilized differently. In terms of the WAIS-R the results raise the hypothesis that this may characterize experts also with neuropsychological instruments. Experts differed also in making more inferences from Response, although Responses have a more marginal role in the RCS than Variables.

There appeared a specific inference type that was associated with the use of the RCS, namely, predictions. While reading the anamnesis, participants sometimes formed predictions as to what the RCS-materials might look like. For example, when anamnesis includes information about interpersonal problems, based on that it would be possible to predict values of certain RCS-variables concerning interpersonal functioning. This phenomenon appeared only in protocols of the expert group. It was done eleven times by five experts (32A, 36A, 38A, 39A, 35B). Although the number of predictions is small, the fact that they appeared in the expert group only shows that this phenomenon characterizes expertise in using the RCS.

3.1.3 Integration

While participants used items and formed inferences related to them, they simultaneously connected information from different sources. An integration episode was identified when the participant, while verbalizing items from one information source, moved to verbalize items from another information source, and combined information from these two sources. This between-source integration combined the RCS-based information with anamnesis, the WAIS-R, or both.

The total number of integration episodes using the RCS was 271. The analyses of these episodes started by examining whether there were differences in the overall number of integration episodes. It appeared that experts executed integration significantly more than novices and intermediates (expert group M=7.3 and SD=3.2; novice group M=4.8 and SD=3.3; intermediate group M=5.1 and SD=2.9).

The nature of these differences was deepened by studying how often integration with anamnesis and integration with the WAIS-R occurred. The RCS was integrated 283 times with anamnesis and 32 times with the WAIS-R. There were only a few instances where all three had been integrated, and thus this phenomenon was not included in analyses. The results of statistical analyses in Table 6 show that there are no differences in how often the RCS was integrated, either with the anamnesis or the WAIS-R, by differently experienced participants. Thus, integration does characterize experts' assessment on a general level, although this does not come forth when we study integration of specific information sources.

TABLE 6   The Number of Different Types of Between-Source Integration Episodes

Type of integration	N		I		E		F/Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
RCS & anamnesis	4.3	3.2	4.4	2.3	6.4	2.9	2.33[a]	2,25	.118	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
RCS & WAIS-R	0.5	0.6	0.8	1.2	0.5	0.6	0.47[c]	2	.805	N<I[d]	ns
										N<E[d]	ns
										I<E[d]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test.
Statistical significance: **p<.01, *p<.05.

Items and inferences. The way to further deepen the understanding of integration was to analyze the use of items and inferences during integration episodes. The results in Table 7 show that there were no differences with regard to Variables, but experts used significantly more Responses than intermediates. With regard to inferences there were no differences. Since integration episodes had items from two or three information sources, during these episodes it was un-

clear what the inferences were connected to, and thus only the number of infer-ences was analyzed.

TABLE 7   The Number of RCS-Materials and Inferences in Between-Source Integration Episodes

Variable	N M	N SD	I M	I SD	E M	E SD	F/Chi	df	Sig	Pw	Sig
RCS-item	7.3	6.6	7.7	5.1	9.8	4.9	0.19[a]	2,25	.828	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Variable	6.7	6.6	7.3	5.3	8.3	4.9	1.27[c]	2	.545	N<I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Response	0.6	1.0	0.4	0.5	1.6	2.1	5.85[c]	2	.048	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	.017*
Inference	14.9	24.3	14.6	13.3	18.5	18.5	1.34[c]	2	.513	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test.
Statistical significance: **p<.01, *p<.05.

In the beginning of this chapter it was presented how similar amounts of items were used by different groups (Table 4). The above results are similar with the exception that during integration experts used more Responses. Since the cen-tral elements of the RCS are Variables and Responses have only a marginal role, it is difficult to see what could be the reason behind this.   Additionally, earlier Table 5 presented results according to which experts made more inferences from Variables and Responses. Higher amount of inferences however did not characterize experts during between-source integration, although one would have expected this.

3.1.4 Errors

So far it has been examined how differently experienced participants used items from different sources, made inferences connected to them, and integrated dif-ferent information sources. With respect to inferences and integration there ap-peared errors, which are analyzed in this section.

Errors in inferences. It was noticed earlier, that the total number of infer-ences made connected to the RCS-items was 2565, and of these 2294 were con-nected to Variables and 271 to Responses. At times, when participant verbalized an inference from a Variable, it was possible to see that the inference was incor-rect. There were in total 580 errors in inferences made from Variables, which

were divided into two types, numerical and conceptual errors. The amount of numerical errors was 170 and the amount of conceptual errors 410. These were statistically analyzed and the results are shown in Table 8. Experts made significantly fewer numerical errors than intermediates, but there was no difference between experts and novices. With regard to conceptual errors experts made significantly fewer errors than participants in both the intermediate and novice groups.

TABLE 8   The Number of Different Types of Errors in Inferences Connected to Variables

Type of Error	N M	N SD	I M	I SD	E M	E SD	F/Chi	df	Sig	Pw	Sig
Numerical Error	1.6	2.4	1.9	1.5	0.6	0.9	8.31[a]	2	.012	N<I[b]	ns
										N>E[b]	ns
										I>E[b]	.005**
Conceptual Error	4.8	3.5	4.3	3.2	1.5	1.6	10.41[c]	2	.003	N<I[d]	ns
										N<E[d]	.035*
										I<E[d]	.000**

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test. [c]F test. [d]Bonferroni test.
Statistical significance: **$p<.01$, *$p<.05$.

Thus, a low number of errors characterizes expertise. Experts are more able to understand the conceptual contents of their instrument. These findings are even more important against the earlier finding that experts often produced more inferences (Table 5), and thus for them the relative amount of errors is very low.

Errors in integration. All the between-source integration episodes were also divided into correct and incorrect ones, using the decision-tree presented in section 2.5.4. The number of integration episodes was 271, which composed of 165 correct episodes and 106 episodes that included errors. The results concerning the total number of integration episodes that included errors are shown in Table 9. According to it, significantly fewer incorrect integration episodes appeared in the expert group than in the intermediate and novice groups.

Errors in integration were analyzed further by dividing the between-source integration episodes in terms of whether the RCS was integrated with the anamnesis or with the WAIS-R. Out of the 283 episodes integrating the RCS and the anamnesis 91 were incorrect. Respectively, out of the 32 episodes integrating the RCS with the WAIS-R 20 were incorrect. Again, the number of episodes integrating all three sources was extremely low and hence they were not included in the analyses. When errors in integrating the RCS and anamnesis were analyzed statistically, the results in Table 9 show that participants in the expert group executed significantly fewer incorrect integration episodes than participants in other groups. In terms of integrating the RCS with the WAIS-R there were no differences.

TABLE 9   The Number of Incorrect Between-Source Integration Episodes and Their Sub-Types

Type of Episode	N M	N SD	I M	I SD	E M	E SD	Chi	df	Sig	Pw	Sig
All episodes	2.8	2.3	2.6	2.4	0.5	0.9	13.74[a]	2	.000	N>I[b]	ns
										N>E[b]	.000**
										I>E[b]	.002**
Sub-Types											
RCS & anamnesis	2.5	2.3	2.1	1.9	0.4	0.7	14.32[a]	2	.000	N>I[b]	ns
										N>E[b]	.000**
										I>E[b]	.003**
RCS & WAIS-R	0.4	0.5	0.6	1.2	0.1	0.3	4.19[a]	2	.129	N<I[b]	ns
										N>E[b]	ns
										I>E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test.
Statistical significance: **p<.01, *p<.05.

The above results can be contrasted with some earlier results, according to which experts produced more between-source integration episodes and more integration episodes where the RCS and anamnesis were integrated. Even though experts generated more integration episodes, they still generated significantly fewer incorrect episodes. This is also related to experts' fewer errors with Variables, since errors in integration is related to understanding or knowing the conceptual contents of the materials.

Between-source integration and errors were still further analyzed in terms of how different RCS-materials were used in correct and incorrect integration episodes. In these episodes there were a total of 466 RCS-items. Of these 418 were Variables, and of these 250 were used in correct integration episodes and 169 in incorrect episodes. Similarly, of the 47 Responses used in these episodes 42 were used in correct and 5 in incorrect episodes. As shown in Table 10, the expert group differed significantly from other groups in using fewer Variables in incorrect between-source integration episodes. With Responses the base rates were too low to allow conclusions.

Related to items many inferences were made during incorrect between-source integration episodes, and also the number of these was analyzed. The number of inferences related to the Variables in all between-source integration episodes was 899, of which 482 appeared in correct episodes and 417 in incorrect episodes. As noticed earlier, during integration episodes it was more difficult to see the item that was connected to an inference, but it was still possible to analyze the number of inferences. The results of these analyses are shown in Table 10, according to which, in incorrect integration episodes the expert group had significantly fewer inferences than other groups. There were no differences between intermediates and novices with this respect.

TABLE 10   The Number of Different Types of RCS-Materials and Inferences in Incorrect Between-Source Integration Episodes

Types of Items and inferences	N M	N SD	I M	I SD	E M	E SD	Chi	df	Sig	Pw	Sig
RCS-item	4.8	5.2	4.4	4.7	0.5	1.1	11.76[a]	2	.001	N>I[b]	ns
										N>E[b]	.003**
										I>E[b]	.002**
Variable	4.5	5.3	4.4	4.7	0.5	1.0	13.93[a]	2	.000	N>I[b]	ns
										N>E[b]	.000**
										I>E[b]	.001**
Response	0.3	0.4	0.0	0.0	0.1	0.2	7.94[a]	2	.020	N>I[b]	.023*[e]
										N>E[b]	ns[e]
										I<E[b]	ns[e]
Inference	12.2	21.3	9.4	10.6	1.8	4.6	9.69[a]	2	.006	N>I[b]	ns
										N>E[b]	.005**
										I>E[b]	.007**

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise. [a]Kruskal-Wallis test. [b]Mann-Whitney test. [e]Too low base rates to justify conclusions. Statistical significance: **p<.01, *p<.05.

Earlier Table 7 presented analyses according to which experts did not generate more Variables or inferences during between-source integration, but above they were shown to produce fewer Variables and inferences when the integration was incorrect. This may mean that experts are more aware about what they know and what not, and they may regulate their use of materials by using less data when they are less sure.

3.2   The Description Phase

After participants completed the Free Assessment Phase, they were asked to give a description of the patient's personality. These descriptions were divided into more simple elements, personality characteristics. In 59 descriptions there were totally 663 characteristics, and thus an average number of personality characteristics in one description was 11.2. This section presents analyses of how groups differed in characterizing personality. The first aim was to examine whether there were differences in the number of given personality characteristics, and according to statistical analyses with this regard there were no significant differences between groups (Table 11).

TABLE 11   The Number of Personality Characteristics

Variable	N		I		E		F	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Personality characteristic	11.2	5.4	10.4	5.5	13.3	5.7	0.97[a]	2,25	.392	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test.
Statistical significance: **p<.01, *p<.05.

Personality characteristics were analyzed with two respects. The first analysis comprised analyzing group differences in terms of the contents of characteristics, and the second one examined the structure of characteristics.

3.2.1 The Content of Personality Characteristics

Personality characteristics were first classified into five main classes: Internal, External, Diagnostic, Treatment, and Other. Respectively, these classes contain characteristics about patient's internal psychological properties, external circumstances, diagnostic issues, treatment issues, and other miscellaneous properties. Internal characteristics were given clearly more often than other types of characteristics, which seems natural given that the task was to assess personality.

The results in Table 12 show that there were no differences with any of these classes. With Diagnostic characteristics the frequencies were too low to justify conclusions. Thus, differently experienced participants produce similar amounts of these types of characteristics, and in this sense characterize personality similarly.

TABLE 12   The Number of Personality Characteristics in the Five Main Classes

Class	N		I		E		F/Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Internal	8.4	4.7	7.7	4.5	10.8	4.7	1.63[a]	2,25	.216	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
External	1.1	1.0	1.3	0.9	1.3	1.4	0.24[a]	2,25	.791	N<I[b]	ns
										N<E[b]	ns
										I>E[b]	ns
Diagnostic	0.1	0.2	0.2	0.4	0.1	0.2	0.80[c]	2	.664	N<I[d]	ns[e]
										N<E[d]	ns[e]
										I>E[d]	ns[e]
Treatment	0.4	0.8	0.3	0.5	0.6	0.8	0.28[c]	2	.875	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Other	0.7	0.8	0.3	0.7	0.4	0.5	3.32[c]	2	.190	N>I[d]	ns
										N>E[d]	ns
										I<E[d]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test. [e]Too low base rates to justify conclusions.
Statistical significance: **$p<.01$, *$p<.05$.

Internal characteristics were further subcategorized into four classes: Experience, Coping, Cognitive, Neuropsychology, Self, Motivation and General. The characteristics that were given most often were about patient's experiences, ways of coping, and general personality properties. The analyses in terms of these classes are shown in Table 13. The only difference was found in terms of Coping characteristics, where experts produced significantly more of these than intermediates. Differently experienced participants behaved differently with all other respects. With the class Motivation the base rates were too low to allow any firm conclusions.

TABLE 13   The Number of Personality Characteristics in the Sub-Classes of Internal Personality Characteristics

Sub-Class	N		I		E		F/Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Experience	2.0	1.8	2.1	2.3	3.3	2.5	1.82[a]	2,26	.182	N<I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Coping	2.3	2.5	1.9	1.2	3.2	2.0	4.14[c]	2	.125	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	.022*
Cognitive	1.1	1.1	1.2	1.7	2.4	1.7	4.83[c]	2	.087	N<I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Neuropsychology	0.1	0.3	0.3	0.6	0.4	0.6	1.86[c]	2	.414	N<I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Self	0.5	0.7	0.2	0.5	0.2	0.4	4.23[c]	2	.116	N<I[d]	ns
										I<E[d]	ns
										I<E[d]	ns
Motivation	0.0	0.0	0.1	0.5	0.2	0.5	2.04[c]	2	.598	N<I[d]	ns[e]
										N<E[d]	ns[e]
										I<E[d]	ns[e]
General	2.2	1.5	1.7	1.7	1.1	1.1	2.63[a]	2,26	.092	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test. [e]Too low base rates to justify conclusions.
Statistical significance: **p<.01, *p<.05.

Some of the above sub-classes of internal characteristics are more related to Clusters in the RCS. Especially, Cognitive characteristics are closely related to so called Cognitive Triad (Information Processing, Mediation, and Ideation), Self characteristics to Self Perception, and perhaps Coping characteristics are more close to Controls Cluster. Additionally, Neuropsychological characteristics are closely related to the WAIS-R. Thus, there are perhaps no differences in utilizing the Cognitive Triad, Self Perception, or the WAIS-R, but experts may utilize more the Controls Cluster.

In the Free Assessment Phase groups were different in terms of inferences, integration and errors. It is worth noting that these differences in the assessment process did not produce differences to the contents of personality characteristics. Especially, although intermediates and novices made significantly more errors than experts, this seems not to be reflected in the contents of characteristics.

3.2.2 The Structure of Personality Characteristics

There were in total 491 internal personality characteristics, the structures of which were analyzed into more elementary structures described in section 2.5.6. These structures were grouped into terms (e.g., *depression*), predicates (e.g., *depressed*), modifiers (e.g., *very*), negation (e.g., *not*), conjunction (*and*), implication (*x causes y*), and contexts (e.g., *those traumatic events that happened to him*). The internal characteristics included in total 2294 of these structures.

Table 14 shows how there were many structural differences in the personality characteristics given by different groups. The key result was that the expert group differed significantly with many structures. It had more predicates (P), modifiers (M), negation (~), conjunction (&) and implication (→) than both other groups. At the same time there were no differences between intermediates and novices in any respect.

TABLE 14 The Number of Different Structure Types in Internal Personality Characteristics

Type of Structure	N M	N SD	I M	I SD	E M	E SD	F/Chi	df	Sig	Pw	Sig
Term (i)	8.9	7.7	9.2	6.5	15.3	7.0	3.33[a]	2,25	.052	N<I[b]	ns
										N<E[b]	ns
										N<E[b]	ns
Predicate (P)	11.3	8.4	12.5	9.5	21.8	8.2	7.17[a]	2,25	.003	N<I[b]	ns
										N<E[b]	.006**
										I<E[b]	.017*
Modifier (M)	2.9	3.7	2.8	2.5	7.3	5.1	9.12[c]	2	.008	N>I[d]	ns
										N<E[d]	.011*
										I<E[d]	.008**
Negation (~)	1.0	1.3	1.2	1.3	3.0	2.7	9.04[c]	2	.007	N<I[d]	ns
										N<E[d]	.007**
										I<E[d]	.015*
Conjunction (&)	2.2	3.6	2.7	2.7	5.4	2.6	9.03[c]	2	.007	N<I[d]	ns
										N<E[d]	.007**
										I<E[d]	.016*
Implication (→)	1.7	2.2	1.9	1.7	3.7	2.3	7.94[c]	2	.014	N<I[d]	ns
										N<E[d]	.012*
										I<E[d]	.028*
Context (c)	0.6	1.2	0.7	0.9	1.6	2.6	3.38[c]	2	.187	N<I[d]	ns
										N<E[d]	ns
										I<E[d]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test.
Statistical significance: **$p<.01$, *$p<.05$.

These results can be compared to some of the earlier findings. First, experts were found to generate a lower number of incorrect between-source integration episodes. It is possible that this is related to their ability to use more Conjunction and Implication – structures that are fundamentally about combining different elements in personality characteristics. Second, experts made fewer numerical errors, which in turn may be related to their ability to use more Modifiers than other participants. Generally, it seems that differences during the Free Assessment Phase have a large effect on the structures of personality characteristics, but not on their contents.

3.3 The Inquiry Phase

After participants had given their view on a patient's personality, the examiner went through the given personality characteristics, one by one, and asked which items in the stimulus materials the characteristics were based on. The analysis of subjects' responses provided a view on how they ground their views and the following sections show the results from analyses of this phase. There were in total 535 grounds, and there were no statistically significant differences between the groups in the number of grounds (Table 15).

TABLE 15   The Number of Grounds

Variable	N		I		E		Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Ground	10.9	5.0	8.7	3.6	9.4	3.7	0.80[a]	2	.683	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test.
Statistical significance: **p<.01, *p<.05.

The number of grounds is naturally related to the number of characteristics that were given (Table 11), since in the Inquiry Phase the examiner asked about each personality characteristic on what in the materials it was based on. In the rest of this section grounds are analyzed in three respects: 1) What kinds of grounds appeared, 2) what those grounds included, and 3) errors in grounds, in terms of the RCS.

3.3.1 Two Types of Grounds

Grounds were first classified in terms of what materials appeared in them. The simple grounds were those where only one source appeared, that is, where only one of the three sources was used. The largest class was that where only RCS-

variables appeared, and the smallest class was grounds utilizing only the WAIS-R or the Responses.

Table 16 shows the number of these respective grounds in different groups. The only difference was that, compared to experts, novices had more grounds that used only anamnesis. With the WAIS-R and Responses the base rates were so low that it was not possible to make conclusions.

TABLE 16   The Number of Different Types of Simple Grounds

Type	N M	N SD	I M	I SD	E M	E SD	Chi	df	Sig	Pw	Sig
Only anamnesis	2.3	2.1	1.8	2.2	0.6	0.9	5.61^a	2	.056	N>I^b N>E^b I>E^b	ns .027* ns
Only WAIS-R	0.1	0.3	0.1	0.2	0.1	0.3	0.86^a	2	.648	N>I^b N>E^b I>E^b	ns^e ns^e ns^e
Only RCS	4.8	5.0	4.5	4.1	4.9	2.7	0.75^a	2	.692	N>I^b N<E^b I<E^b	ns ns ns
Only Variable	4.3	5.1	4.3	4.0	3.8	2.9	0.55^a	2	.770	N>I^b N<E^b I<E^b	ns ns ns
Only Response	0.1	0.3	0.0	0.0	0.4	0.8	5.81^a	2	.034	N>I^b N<E^b I<E^b	ns^e ns^e ns^e

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
^{a}Kruskal-Wallis test. ^{b}Mann-Whitney test. ^{e}Too low base rates to justify conclusions.
Statistical significance: **p<.01, *p<.05.

In complex grounds elements from materials were integrated in different ways. In terms of the RCS, integration in grounds consisted of two phenomena: between-source integration, where the RCS was integrated with either anamnesis or the WAIS-R, and within-source integration, where elements within the RCS were integrated. Of all 535 grounds the number of grounds including between-source integration was 168, and of these 160 included the RCS. Thus, there were 8 grounds integrating the anamnesis with the WAIS-R, without the RCS. Respectively, the number of grounds including within-source integration was 286. The statistics in Table 17 show that even though experts gave more between-source integration grounds, there were no statistically significant differences in any respect. Exactly the same was true of the within-source integration grounds.

TABLE 17   The Number of Different Types of Grounds with Integration

Type of Ground	N		I		E		F/Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Between-Source Integration											
RCS & anamnesis	2.5	2.2	1.9	1.7	2.5	1.8	0.60a	2,25	.558	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
RCS & WAIS-R	0.3	0.6	0.1	0.3	0.6	0.9	2.78c	2	.279	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
RCS & WAIS-R & anamnesis	0.3	0.6	0.2	0.5	0.5	0.9	0.65c	2	.761	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
All types	3.1	2.0	2.2	1.7	3.4	2.5	1.26a	2,25	.301	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Within-Source Integration											
RCS & RCS	5.0	3.5	3.8	2.3	6.5	3.2	1.12a	2,25	.341	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test.
Statistical significance: **$p<.01$, *$p<.05$.

The above results correspond to those in the Free Assessment Phase in that also there no differences in terms of the number of different kinds of between-source integration episodes appeared (Table 6).

3.3.2 Items

The simplest unit in grounds is an item, which is an element from the materials. For example, an item from the RCS would be any Variable or Response. There were in total 1434 items, of which 226 were about anamnesis, 45 about the WAIS-R, and 1163 about the RCS. Of all RCS-items 1077 were Variables and 86 were Responses. According to the results in Table 18 the only difference was that the expert group had significantly more Responses than the intermediate group. With regard to the anamnesis, WAIS-R, and RCS-variables, the partici-pants used items from these sources similarly.

TABLE 18   The Number of Items from Different Sources in Grounds

Source	N		I		E		F	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Anamnesis	5.1	2.7	4.0	3.1	3.3	2.4	1.26[a]	2,25	.302	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns
WAIS-R	0.9	1.1	0.6	0.8	1.0	1.3	0.75[a]	2,25	.482	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
RCS	17.6	11.4	16.4	10.6	27.7	14.9	2.82[a]	2,25	.079	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Variable	16.3	10.3	15.9	10.7	24.9	15.4	1.84[a]	2,25	.180	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Response	1.3	2.3	0.5	0.7	2.8	3.1	3.88[a]	2,25	.034	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	.034*

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni-test.
Statistical significance: **p<.01, *p<.05.

These results converge with some of the earlier findings: Experts did not use more WAIS-R, RCS-items, or Variables (Table 4), but a similar finding was obtained in between-source integration episodes, where experts also used more Responses than intermediates (Table 7). However, in the Free Assessment Phase experts did not generally use more Responses (Table 4), but in giving grounds they did use more of them than intermediates.

Items in grounds with integration. Grounds that included between-source integration combined the RCS with the anamnesis, with the WAIS-R, or with both of these. From these grounds the number of RCS-materials was analyzed. According to the results in Table 19, although all items appeared more in the expert group, the only statistically significant difference was that experts had more Responses than both intermediates and novices, but this conclusion is not reliable because of too low base rates. The result that there are no significant differences in terms of RCS-items and Variables during between-source integration is similar to findings from between-source integration in the Free Assessment Phase (Table 7).

Table 19 also shows analyses concerning the use of the RCS during within-source integration. Experts used significantly more RCS-items than intermediates, and experts used significantly more Responses than both intermediates and novices. In terms of Variables the mean in the expert group was clearly higher but not statistically significant.

TABLE 19   The Number of Different Types of RCS-Materials in Grounds with Different Types of Integration

Type of Integration & Type of Item	N M	N SD	I M	I SD	E M	E SD	F/Chi	df	Sig	Pw	Sig
Between-Source Integration											
RCS-item	8.9	7.1	5.8	5.7	15.7	8.7	0.53[a]	2,26	.099	N>I[b]	ns
										N<E[b]	ns
										I<E[b]	ns
Variable	7.8	6.9	5.3	5.5	11.1	9.2	3.82[c]	2	.190	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Response	1.1	2.4	0.4	0.6	4.6	5.4	9.36[c]	2	.006	N>I[d]	ns[e]
										N<E[d]	.039*[e]
										I<E[d]	.002**[e]
Within-Source Integration											
RCS-item	14.6	9.8	13.4	10.0	25.3	14.9	4.24[c]	2	.115	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	.045*
Variable	13.4	8.7	13.0	9.9	23.2	15.2	3.28[c]	2	.198	N>I[d]	ns
										N<E[d]	ns
										I<E[d]	ns
Response	1.1	2.5	0.5	0.7	2.1	2.7	8.50[c]	2	.010	N>I[d]	ns
										N<E[d]	.034*
										I<E[d]	.005**

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]F test. [b]Bonferroni test. [c]Kruskal-Wallis test. [d]Mann-Whitney test. [e]Too low base rates to justify conclusions.
Statistical significance: **$p<.01$, *$p<.05$.

3.3.3 Errors

Items. When participants gave grounds for personality characteristics there appeared two kinds of errors with the RCS. The first type represents an inappropriate use of the RCS, where a participant uses it to argue for a personality characteristic for which it is not possible to argue with the RCS. From the total of 535 grounds 70 were such where the RCS was inappropriately used. Table 20 shows that there were no differences between groups in the number of inappropriately used RCS-items in these grounds.

The second kind of error was to use the Variables incorrectly, that is, to reason with a wrong Variable. It was appropriate to use Variables in 465 grounds, and of these 124 contained incorrect uses of Variables. According to Table 20 the expert group differed significantly from other groups in using fewer Variables incorrectly. This verifies an earlier finding that experts make fewer conceptual errors than participants in other groups (Table 8).

TABLE 20   The Number of Different Types of Erroneous Uses of Variables in Grounds

Type of Use	N		I		E		Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Inappropriate	3.3	3.6	3.4	3.8	2.6	4.0	0.65[a]	2	.735	N<I[b]	ns
										N>E[b]	ns
										I>E[b]	ns
Incorrect	6.2	4.4	5.5	5.3	1.5	2.5	8.71[a]	2	.009	N>I[b]	ns
										N>E[b]	.009**
										I>E[b]	.012*

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test.
Statistical significance: **p<.01, *p<.05.

Between-source integration. Errors with the RCS were also analyzed as they appeared in grounds with between-source integration. Inappropriate and incorrect ways to use the RCS were analyzed separately with regard to different ways to integrate materials. The only statistically reliable results were obtained from analyses where the RCS and anamnesis were integrated because mostly the base rates were too low. In these cases, both inappropriate and incorrect use of Variables occurred as often in all groups.

TABLE 21   The Number of Different Types of Uses of the RCS in Different Types of Between-Source Integration Grounds

Type of Use and Type of Integration	N		I		E		Chi	df	Sig	Pw	Sig
	M	SD	M	SD	M	SD					
Inappropriate Use											
RCS & anamnesis	0.9	1.2	0.6	0.8	0.3	0.7	4.26[a]	2	.119	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns
RCS & WAIS-R	0.1	0.3	0.1	0.2	0.0	0.0	2.81[a]	2	.261	N>I[b]	ns[c]
										N>E[b]	ns[c]
										I>E[b]	ns[c]
All Three Sources	0.0	0.0	0.0	0.0	0.1	0.3	3.74[a]	2	.312	N>I[b]	ns[c]
										N<E[b]	ns[c]
										I<E[b]	ns[c]
Incorrect Use											
RCS & anamnesis	1.0	1.2	0.8	1.1	0.5	0.9	1.67[a]	2	.446	N>I[b]	ns
										N>E[b]	ns
										I>E[b]	ns
RCS & WAIS-R	0.1	0.3	0.0	0.0	0.3	0.4	4.75[a]	2	.078	N>I[b]	ns[c]
										N<E[b]	ns[c]
										I<E[b]	ns[c]

Table 21 continues

	N		I		E						
	M	SD	M	SD	M	SD	Chi	df	Sig	Pw	Sig
All Three Sources	0.2	0.6	0.1	0.3	0.1	0.2	2.19[a]	2	.375	N>I[b]	ns[c]
										N>E[b]	ns[c]
										I>E[b]	ns[c]

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test. [c]Too low base rates to justify conclusions.
Statistical significance: **$p<.01$, *$p<.05$.

Within-source integration. Also within-source integration was analyzed with re-gard to errors (Table 22). With regard to the inappropriate uses of the RCS there were no differences. In terms of incorrect uses of Variables the expert group differed significantly from other groups in using fewer Variables incorrectly. This latter finding is again similar to the earlier result that expert made fewer conceptual errors (Table 8).

TABLE 22   The Number of Different Types of Erroneous Uses of the RCS in Within-Source Integration Grounds

	N		I		E						
Type of Use	M	SD	M	SD	M	SD	Chi	df	Sig	Pw	Sig
Inappropriate	2.3	2.6	2.5	3.3	1.8	3.6	1.35[a]	2	.521[a]	N<I[b]	ns
										N>E[b]	ns
										I>E[b]	ns
Incorrect	4.7	3.8	4.7	5.3	1.2	2.2	7.43[a]	2	.019	N>I[b]	ns
										N>E[b]	.023*
										I>E[b]	.015*

Note. N = Novice group, I = Intermediate group, E = Expert group, Pw = Pairwise.
[a]Kruskal-Wallis test. [b]Mann-Whitney test.
Statistical significance: **$p<.01$, *$p<.05$.

In analyzing errors in grounds there occurred a consistent finding, according to which there are no differences in how often inappropriate use of Variables oc-curs in different skill groups. This indicates that already at the novice level participants understand quite well when it is not possible to give grounds with the RCS, or vice versa, none of the groups understand this well. However, it is also possible that there are some presently unknown conceptual issues hidden in the category of inappropriate use of Variables.

4   DISCUSSION

4.1   Aims of the Study

The primary aim of this study has been to explore how clinical psychologists, who are differently experienced in using the RCS, use it during clinical personality assessment. This has meant to study the RCS-based assessment process on the different clinician-level areas (Table 1). This aim has been accomplished by applying methodology and design from the expertise research to clinical personality assessment. The results are provisionally conceptualized using concepts from the expertise research, e.g., the nature of knowledge, integration, forward and backward reasoning, errors, and skill development. Concerning expectations, earlier studies on expertise in using the Rorschach indicate that there would be no differences between groups, while expertise research from other problem-solving areas (like medical diagnostics) do show differences. This chapter summarizes the results, explores the above issues, discusses results in terms of the RCS, and discusses questions concerning methodology and design. In the end, some ideas regarding teaching and future research are presented.

4.2   Summary of Findings

Before discussing the findings, this section summarizes them dividing the findings where differences between groups were found from those where no differences came forth.

Differences between groups. The only differences detected were those between experts versus less experienced participants. Experts differed with regard to six different phenomena. The first difference was their better ability to use Variables. They made more inferences from Variables (Table 5), made fewer numerical errors and conceptual errors (Table 8), used fewer Variables incorrectly in grounds (Table 20), and used fewer Variables incorrectly while integrating Variables with other Variables in grounds (Table 22).

The second phenomenon with clear differences was experts' ability to execute between-source integration. In addition to integrating more, experts used more Responses in integration (Table 7), executed fewer incorrect integration episodes generally and more specifically between the RCS and anamnesis (Table 9), used fewer Variables during incorrect integration (Table 10), and formed fewer inferences during incorrect integration (Table 10).

The third area of difference was that experts often used more Responses differently: They made more inferences from Responses (Table 5), used more Responses in between-source integration (Table 7), used more Responses in within-source integration (Table 19), and used more Responses in grounds (Table 18). Fourth, experts were also different in terms of integrating Variables with other Variables: They used more RCS-items during within-source integration (Table 19) and integrated fewer Variables incorrectly (Table 22). Fifth, experts made inferences differently: they made more inferences from both Variables and Responses (Table 5), made fewer numerical and conceptual errors in inferences (Table 8) and made less inferences during incorrect integration between information sources (Table 10).

Sixth, experts also differed in terms of the nature of personality characteristics. In terms of the contents of personality characteristics the only difference was that experts produced more characteristics describing Coping (Table 13). The major differences were in the structures of personality characteristics. The characteristics given by experts included more Predicates, Modifiers, Negation, Conjunction and Implication (Table 14).

Finally, although the aim was to analyze only the use of the RCS, in can be noticed that in terms of the other two information sources there were three differences between groups: experts made more inferences than novices from the WAIS-R (Table 5); experts had fewer incorrect between-source integration episodes using both of these sources (Table 9); and experts had a lower number of grounds using only anamnesis (Table 16).

No differences between groups. There were no differences whatsoever between intermediates and novices. In addition, on some of the findings all the three groups were similar. This occurred with five different phenomena. First, there were no differences in the overall number of used Variables when correctness was not taken into consideration (Tables 4 and 18). Second, when correctness was not taken into consideration, the number of integration episodes between specific sub-types of information sources (RCS and anamnesis; RCS and WAIS-R; RCS and anamnesis and WAIS-R) was similar (Tables 6 and 17). Third, when correctness was not taken into consideration, the number of RCS-items and Variables was similar with regard to between-source integration and within-source integration (Tables 7 and 19). Fourth, there were no differences with regard to the inappropriate use of Variables (Tables 20, 21 and 22), and fifth, also no difference in the number of incorrectly used Variables in between-source integration in grounds was found (Table 21). Sixth, there were no differences in the number of episodes integrating the RCS with the WAIS-R, regardless of whether the episodes included errors or not (Tables 6, 9 and 17). Seventh,

there were no differences in the number of inferences in between-source integration (Table 7). Eighth, concerning personality characteristics, there were no differences in the number of given characteristics (Table 11); in the number of Internal, External, Treatment or Other characteristics (Table 12); in the number of all other sub-classes of Internal characteristics except Coping (Table 13; with the class Motivation the base rates were too low to justify conclusions); and in the number of two structures of characteristics, Terms and Contexts (Table 14). Ninth, there were also no differences in the number of given grounds (Table 15), although this was directly related to the number of given personality characteristics. Tenth, there were no differences in the number of simple grounds, using only the RCS (Table 16).

Finally, in can be noticed that in terms of the anamnesis and the WAIS-R there were no differences in the following respects: the number of used items from these sources (Tables 4 and 18); inferences made from anamnesis and the WAIS-R (Table 5); the number of correct between-source integration episodes using these sources (Tables 6 and 17); and the number of incorrect between-source integration episodes using the WAIS-R (Table 9).

4.3 Expertise in the RCS-Based Personality Assessment

4.3.1 Cognitive Interpretation of Results

This section discusses the results by comparing them to some essential findings and concepts obtained from the research on medical diagnostic expertise. However, these comparisons must be understood as only preliminary discussions. The reason for this is twofold. First, concepts rising from expertise research in medical diagnostics are domain sensitive in that they have evolved from very specific empirical findings. Therefore, they are not easily generalized into any other field of study. Second, results of this thesis are also domain sensitive and are therefore not readily conceptualized by using concepts from other fields. For these two reasons the discussions in this section at best serve to give only preliminary conceptual forms for findings.

Extent of knowledge. Expertise research has repeatedly shown how experts possess more knowledge in their task environment than do less experienced individuals (e.g., Chi, 2006; Horn & Masunaga, 2006). Chi (2006) gives two examples of this that may be relevant here: experts have more knowledge, and their knowledge structures or representations are more detailed. The results of this thesis indicate that experts in personality assessment with the RCS may be superior in both of these respects. In terms of sheer amount of knowledge, the fact that experts made fewer errors (in inferences from Variables, in integration, in using Variables in grounds, and in integrating Variables with other Variables in grounds) may witness that their knowledge base is larger. Although expertise research defines the amount of knowledge a little bit differ-

ently, using different kinds of ways to measure the amount of bits of knowledge, the findings concerning errors do point to the same direction.

In terms of the more detailed or complete nature of experts' knowledge structures, the findings concerning the larger amount of many different kinds of structural elements in experts' descriptions of personality seem to be close to this idea. The personality characteristics that experts gave included more Predicates, Modifiers, Negation, Conjunction and Implication, and in this sense their knowledge structures are much more detailed.

Scientific and clinical knowledge. In contexts medical diagnostics knowledge has been further divided into subtypes of which two are scientific knowledge and clinical knowledge, or deep knowledge and shallow knowledge, respectively. Gilhooly (1996), for example, understands scientific or deep knowledge as referring to knowledge about pathophysiology and causal mechanisms of diseases, while clinical or shallow knowledge refers to clinical knowledge of associations between symptoms and diagnoses based on previously experienced clinical cases. It has been noticed in the domain of physics that experts display the features of problems at the level of principles of physics, while less experienced physicists use features that are more concrete or superficial (e.g., Patel & Arocha, 2001). With respect to the physicians, however, the findings have been the opposite with expert physicians using scientific knowledge less than less experienced physicians (Boshuizen & Schmidt, 1992; Patel & Groen, 1986b).

The results of this thesis do not seem to be easily conceptualized with regard to scientific versus clinical knowledge. A main reason for this is that the categories that emerged from the data do not reflect this kind of differentiation. First, experts' knowledge related to the RCS is different from physicians' knowledge of pathophysiology. These may be analogical in the sense that both represent scientific knowledge, and in this respect such RCS-concepts as thought disorder, introversive problem-solving style or modulation of emotions may come close to scientific expressions in medicine. However, the RCS also includes many concepts that come quite close to some everyday language idioms. Good examples would be concepts such as anger, distress and isolation. Second, what is relevant here is not what parts of the RCS are used, but how those materials are conceptualized. A Variables indicating anger can probably be conceptualized using both scientific and clinical knowledge. To study this one would need to classify all the inferences in the results of this thesis according to whether they represent scientific or clinical knowledge. Further, it is not easy to determine how to make this kind of division in clinical psychology, which is different from medical diagnostics based on natural science. Third, similarly, it seems possible that in clinical psychology anamnestic information can be conceptualized using either scientific or clinical knowledge. Therefore, to conceptualize these results in terms of different kinds of knowledge would take some other kind of differentiation. One example might be a differentiation between RCS-concepts, theoretical psychological concepts, and everyday language psychological concepts. Finally, there however was some information in

the results that comes close to scientific knowledge in medical diagnostics. While scientific knowledge in medical diagnostics means knowledge about causal mechanisms, this is a little bit similar to the result that experts produced personality characteristics that included more descriptions of causal relationships in the materials. But this is still not equivalent in that causality probably has very different meanings in medicine and clinical psychology.

The integration and organization of knowledge. Integration was one of the most essential findings in the results. Experts executed between-source integration more and more correctly, and the same held true in terms of within-source integration. Integration has also been an essential finding in medical expertise research. That research indicates that as the level of expertise increases the elements in the knowledge structures become more organized and integrated. In medical diagnostics integration of knowledge refers to the degree to which concepts are related to one another in different meaningful ways. Evidence for this has been found in, e.g., how expert physicians are able to form clusters of observations with respect to their diagnostic interpretations in a perceptual task (Patel & Arocha, 2001). Novices solve problems based more on concrete features in the given materials, while experts use more derivative features, where the interactions of the concrete entities are integrated to describe the problem situation in a different manner (Chi et al., 1981). The resulting knowledge system provides the expert with a basis for selecting, organizing, representing, manipulating, and interpreting information in the environment (Horn & Masunaga, 2006).

There are several findings in this thesis that may be conceptualized accordingly. First, experts generally integrated more between information sources. Second, they produced personality characteristics that were structurally more complex. While personality characteristics were the result of the RCS-based assessment process, it is possible that the properties of assessment process are reflected in the nature of personality characteristics. That is, more integrated knowledge structures during the assessment process may be the reason behind more complex personality characteristics.

Forward and backward reasoning. Based on empirical studies a difference between forward reasoning and backward reasoning has been drawn (Gilhooly, 1996; Nickerson, 1994; Patel & Groen, 1986a; Patel & Groen, 1991a). In forward reasoning a person works from the given information towards the solution, whereas in backward reasoning the working proceeds from a hypothetical solution to the given information. It has been noted that experts make greater use of forward reasoning while novices tend to reason backwards (Patel & Groen, 1986a). E.g., Patel et al. (Patel & Groen, 1986a; Patel, Groen & Arocha, 1990) used this distinction to interpret results from studies, where they examined expertise in diagnosing acute bacterial endocarditis. They modelled graphically the reasoning process abstracted from verbal protocols, and identified forward reasoning in cases where the participant proceeded from an item in the given materials to forming of a hypotheses – and vice versa for backward reasoning. One example of forward reasoning was reasoning from "puncture wounds and

young unemployed male" to "intravenous drug use"; another example was to reason from "fever" to "bacterial infection" (Patel & Groen, 1990, p. 398).

It is possible that something equivalent to these two types of reasoning can be found in the results of this thesis. In the Free Assessment Phase an example of forward reasoning may be that experts generated more inferences from Variables and Responses. If so, this would fit findings from medical diagnostics, according to which these two types of reasoning are different in terms of errors made during problem-solving. Forward reasoning is more error-prone since the inferences cannot be checked, and it thus necessitates a great deal of knowledge. Similar finding was obtained here. Experts made fewer errors when generating inferences from Variables. Because experts have more knowledge in their domain than do the novices, they can use this knowledge to solve problems by means of forward reasoning (Gilhooly, 1996).

Errors. Errors during the RCS-based personality assessment were one of the key differences between differently experienced groups. This is a very important phenomenon that has directly to do with both clinical activities and training. In medical diagnostics, cognitive mechanisms are the major factor that contributes to medical errors (Patel, Arocha & Zhang, 2005). The interpretation of errors during personality assessment process is probably down to several different kinds of knowledge structures, and through elaborating these in the future it may be possible to trace more exactly the factors behind errors.

The development of skills. One general expectation in this thesis was for there to be monotonically increasing differences in the use of the RCS as a function of increasing experience. The results did not support this idea, since there were no differences between intermediates and novices in any respect. This may imply that there is something special in how expertise in RCS-based personality assessment develops since the skill development may be non-existent for a long time. Although the results are in line with generalizations made from expertise research, according to which it often takes about ten years to gain expertise (Richman, Gobet, Staszewski & Simon, 1996), the non-development in lower experience levels is still counter-intuitive. This finding may suggest that in the intermediate group the development of skills has stopped at a routine level, called arrested development (Ericsson, 2006b). Development above this routine level would necessitate deliberate practice, that is, practice that deliberately intends to enhance the weak areas of performance (Ericsson, 1996; Ericsson, 2006b; Ericsson & Charness, 1997). Arrested development may have been facilitated by the fact that the main experience of intermediate and novice level clinicians consists of everyday work activities. Another interpretation would be that the RCS is so difficult a method that developmentally intermediates are not yet in a position to master it. The more difficult the method, the more it has to be practiced in order to enhance skills.

4.3.2 Expertise and the Rorschach

One main expectation in this study, based on earlier studies using the Ror-
schach, was that experts would not be different from less experienced partici-
pants. However, the results were quite the opposite in many respects. The re-
sults strongly argue in favor of the view that expertise is a real phenomenon in
the RCS-based personality assessment.

There are several large differences between this study and the earlier stud-
ies on the nature of expertise in using the Rorschach, which are commented on
in this section. One essential difference concerns the way levels of experience
have been defined. In this study the level of experience with the RCS was one
year in the novice group, 3-9 years in the intermediate group, and 9-30 years in
the expert group. In earlier studies the experience was categorized in the fol-
lowing ways: 10 years/six years/three years/one semester course/no knowl-
edge at all (Grant, Ives & Ranzoni, 1952); thoroughly trained and experi-
enced/in training (Bialick & Hamlin, 1954); minimum of 5 years/second-year
level of training/an introductory course in administering and scoring (Greb-
stein, 1963); at least 10 years / less than 5 years/a course in administration/no
knowledge at all (Turner, 1966); at least 3 years (Chambers & Hamlin, 1957); 4-9
years (Armitage & Pearl, 1957); mean 8 years (Lisansky, 1956); mean 20.6 years
(Albert, Fox & Kahn, 1980); Diplomates (Filmer-Bennett, 1955); and 5 or more
years (Winslow & Rapersand, 1964). Although these studies were not intended
to examine expertise, they still blur the concept of an expert or experienced par-
ticipant in the sense that there has to be more explicit definition of these con-
cepts. In this respect the definitions in the study by Grant, Ives & Ranzoni
(1952) seem quite adequate, but there was only one participant on each level.
Grebstein (1963) and Turner (1966) had groups that might well give information
about differently experienced participants' use of the Rorschach, but in these
studies the given information was artificially restricted.

A further difference concerns the materials used. The given materials were
often very restricted: contents (Armitage & Pearl, 1957; Bialick & Hamlin, 1954;
Winslow & Rapersand, 1964), structural data (Armitage & Pearl, 1957; Greb-
stein, 1963); and all Rorschach data (Albert, Fox & Kahn, 1980; Armitage &
Pearl, 1957; Chambers & Hamlin, 1957; Filmer-Bennett, 1955; Lisansky, 1956;
Turner, 1966). There may have been good reasons for some of these decisions,
but even though different Rorschach Variables were used in assessing different
issues, personality assessment should utilize all information – also that other
than the Rorschach (Exner, 2003). Another concern is the nature of the used in-
formation. From the present-day perspective, using Rorschach contents as the
basis for decisions would mean to base decisions on something other than valid
information, since the contents most probably do not have sufficient research
base behind them.

Another difference is in the object of study. The focus in the earlier re-
search has been on the nature of judgments or decisions, but in this thesis it is
on the nature of the personality assessment process and freely given characteri-
zations of personality. More specifically, in earlier research judgments have

concerned adjustment (Grant, Ives & Ranzoni, 1952), intelligence (Bialick & Hamlin, 1954; Grebstein, 1963), diagnosis (Armitage & Pearl, 1957; Chambers & Hamlin, 1957), personality (Lisansky, 1956; Turner, 1966), malingering (Albert, Fox & Kahn, 1980), and treatment outcome (Filmer-Bennett, 1955; Winslow & Rapersand, 1964). There is no doubt that the RCS can be used in assessing at least adjustment, personality and treatment outcome, but one should be very skeptical in trying to do these assessments based on contents, Whole Responses or protocols alone.

There are also differences in terms of the Rorschach systems used. Some used the Klopfer-system (Grant, Ives & Ranzoni, 1952; Turner, 1966) and some the Beck-system (Grebstein, 1963), and many of the studies did not report the system at all. There are many differences between the earlier systems (Exner, 1969), bringing complications to comparisons. In one of the studies (Grant, Ives & Ranzoni, 1952) the Klopfer-system was used but one of the participants was trained to use the Herz-system.

In terms of the results of the earlier studies, the experts consistently were not more successful than less experienced judges. This happened with judging adjustment (Grant, Ives & Ranzoni, 1952), intelligence (Bialick & Hamlin, 1954; Grebstein, 1963) or personality descriptions (Turner, 1966). However, in one study they were not worse than actuarial models when judging intelligence (Grebstein, 1963). Sometimes they judged diagnosis accurately (Chambers & Hamlin, 1957), but sometimes did not (Armitage & Pearl (1957); sometimes they were able to correctly judge treatment outcome (Winslow & Rapersand, 1964), but sometimes not (Filmer-Bennett, 1955); sometimes they agreed more when judging sensitivity in human relationships and diagnosis (Lisansky, 1956); and sometimes they were not accurate in detecting malingering (Albert, Fox & Kahn, 1980).

This thesis did not ask participants to judge in terms of fixed set of alternatives, but allowed them to freely create their views. Although in real life clinicians sometimes have to choose from given options in the referral question, they still have to make their assessments using a large amount of information. When large amount of materials was used in this study, the experts did differ in many respects from less experienced participants. Giving participants fixed options is problematic in that the results may not answer the question of how the clinician really understands and uses the materials. This was seen in some very counter-intuitive results of the earlier studies reported above. Participants with no knowledge of the Rorschach fared as well as experienced participants (Bialick & Hamlin, 1954; Grant, Ives & Ranzoni, 1952; Grebstein, 1963; Turner, 1966). It would not seem possible that in a kind of design used in this thesis, where participants have to work using their knowledge structures, such totally unknowledgeable participants could proceed in the assessment process.

4.3.3 Expertise and the RCS

Exner (2003) seems to define four different levels when using the RCS: interpretation of Variables and Responses, interpretation process defined by steps, integration of Variables, and integration of Clusters. The first and second are based on empirical research, and two latter ones are theoretical or clinical views of Exner (2003). The results of this thesis provide information on the first three areas.

The use of the RCS-materials was reflected in the categories of items and inferences. In terms of items, there were no statistical differences in their use, implying that they were utilized similarly. Relevant here is the fact that the means were clearly lower than the number of Variables in the RCS. The numbers of Variables were 38.2 for experts, 28.5 for intermediates and 26.9 for novices, and these numbers indicate that participants did not go through all the structural materials. This deviates from the RCS, where all Variables in the interpretational steps should be gone through. It also tells that participants did not follow the steps in their interpretation process, which violates the idea of stepwise interpretation process. The minimum values in the two cases were the following: 29 (Case A) and 29 (Case B) for experts; 14 (Case A) and 19 (Case B) for intermediates; 6 (Case A) and 14 (Case B) for novices. As these figures show, sometimes clinicians use only a very small part of the Variables.

The above findings express that in no group did the participants utilize all of the information available, whereas they did go through the clinical information in its wholesomeness since they read all of the anamneses. One possible way to interpret these findings is offered by the concept of selectivity, which has been used, for example, in analyzing the thinking of chess players (de Groot, 1965; Saariluoma, 1995). It means the occurrence of selecting a specific information element from among a larger amount of information. Selectivity naturally is bound to the nature of knowledge structures. According to the interpretation of results made earlier experts are superior with regard to the nature of their knowledge compared to less experienced participants. Therefore, selectivity may be different for these groups. It is probable that for many experts it is possible to select material that is relevant, whereas novices use some other criteria in choosing information from the RCS-materials.

In the RCS both Variables and Responses are interpreted, but only with Variables specific meanings are defined. In analyzing the number of inferences the only difference was that experts formed more inferences from Variables than novices. Experts formed 11 inferences per 10 items, but novices only 7. Thus, novices did not interpret every Variable. They probably were not able to, considering their narrower knowledge base.

A further difference was that experts made fewer errors with Variables than both other groups, and experts' inferences were thus qualitatively different by being more in accordance with Exner's (2000) definitions. When the amounts of inferences were examined relative to errors, there were 6.7 errors per 25.0 inferences in the novice group, 6.2 errors per 41.6 inferences in the intermediate group, and 2.1 errors per 53.1 inferences in the expert group. If we suppose that

numerical and conceptual errors concern different inferences (as mostly was the case), 27% of novices' and 15% of intermediates' inferences included errors, but with experts the percentage was only 4%. Experts' knowledge structures are thus almost totally in accordance with Exner's definitions. Additionally, while both intermediates and novices had undergone about 13-14 months of training, it is possible that there is something in the RCS that is too difficult for them. In this sense the RCS might be somewhat simpler, or then the training should be different in kind. It is not conceivable that the majority of clinicians (novices and intermediates most probably comprise the largest group of clinicians) do not completely handle the knowledge that their method includes. A possibility to this effect would be to teach intermediates and novices metacognitive skills to notice instances when their knowledge is insufficient, and use clinical information instead.

In terms of the Responses, during the era prior to Exner's creation of the RCS in 1974, some of the developers of the Rorschach systems had put more weight on content analysis than is being done in the RCS. E.g., Schafer (1954) developed interpretation based largely on the Responses. In their survey Exner and Exner (1972) found that many clinicians who used the Rorschach preferred to use contents in favor of Variables, and to gain more validity, in the RCS the use of Responses was diminished in favor of the Variables (Exner, 2003). Against these facts the results of this study are counter-intuitive in showing that experts used more Responses than intermediates and made more inferences from Responses than novices. It is possible to interpret this using a piece of information that was not reported in Chapter 3. Within the expert group there seemed to be different attitudes towards using Responses. Especially three participants (32, 34, 39) placed much weight on Responses in making psycho-dynamic-like inferences concerning the personality of the patient. At the same time there were two (37, 38) who emphasized Responses in a way that one day may become part of the RCS, that is, the RCS as a neuropsychological kind of instrument (Perry et al., 1996; Perry & Potterat, 1997). The other expert participants (31, 33, 35, 36) used the Responses in some other way. Hence, it may be that within the expert group there are sub-groups with different styles or approaches in using the Responses. However, although Exner (2003) does not give the Responses the same scientific status as he gives the Variables, he does clearly define how clinicians should search which Responses to interpret and which not. In this respect it seems that participants have not internalized these procedures.

Exner (2003) sees that the interpretation of Variables is not a set of fragmented interpretations, but Variables must be understood as being related to each other. Otherwise "the likelihood of error is substantial because this kind of interpretive strategy ignores additional data that will bear directly on whether the deviation based hypothesis should be accepted, rejected, or modified" (Exner, 2003, p. 219). Although the stepwise interpretation procedure is empirically based, Exner's attitude of integrating Variables with Variables lacks scientific status in the sense that there is no empirically based scientific knowledge of the

relationships between specific Variables. Since however the Variables them-selves have a scientific basis, their integration should be based on this knowl-edge. Since intermediates and novices made more errors with Variables, it is also understandable that they were not able to pursue these actions correctly.

There seems to be an enormous bias between the amount of research on the test-level and on the clinician-level. As the above considerations show, there are cognitively speaking many different dimensions in the clinician-level where the use of the RCS is not scientifically based. The conclusion is that developing the method is not enough, but there also has to be empirical research and the-ory development concerning mental activity, cognitive processes and mental contents therein. If before the development of the RCS practicing clinicians did not follow any system faithfully (Exner & Exner, 1972), it may be that without research on clinician-level processes they may not follow the RCS faithfully. As this study shows, no matter how good a test is, it is possible that the use of it brings cognitive factors that endanger the obtained test-level reliability and va-lidity.

4.3.4 Summary

Chapter 1 presented a systematic overview of research on the RCS-based per-sonality assessment, and it was pointed out that most of the research has been done on the test-level, while the clinician-level includes only very few studies. Based on the lack of clinician-level research, Cates (1999) considers the clinician-level of personality assessment to be more art than science, and even goes as far as to state that this clinician-level approach cannot be researched. There is how-ever no reason for such a pessimistic opinion. As this study shows, protocol analysis is a completely appropriate method for gathering data on this matter, and this data can perfectly well be analyzed. Thus, personality assessment can and should be science.

4.4 Evaluation of Methodology

4.4.1 Participants and Groups

The key question concerning the representativeness of participants is the defini-tion of an expert. Some domains, such as sports, music and chess include com-petitions that make it relatively easy to say who is an expert and who is not. In personality assessment this kind of comparison is difficult, and in this study the problem was solved by choosing acknowledged experts from the Society for Personality Assessment and then using the experience (in years) as an addi-tional criterion. The result was a group that numerically (M=20.1 years, SD=7.3) well corresponds to the ten year rule of Richman et al. (1996), according to which it takes about ten years to acquire expertise. A related question is the definition of other groups. The difference between the intermediate and novice

groups seems to be large enough to be taken to represent different levels of experience (intermediate group M=6.2, SD=2.3; novice group M=1, SD=0). Another question concerns the difference between expert and intermediate groups. The large difference in the results between these two groups could be understood as an effect of large differences in experience (M=20.1 versus M=6.2). Although there may be discontinuity between these groups, this most probably does not affect the interpretation of group differences, since what is more crucial is the non-difference between intermediate and novice groups which tells about arrested development. However, a more precise skill developmental view might need additional group or groups that would be located between what was here defined as experts and intermediates.

A question that concerns the representativeness of intermediate and novice groups is that the criterion that was used to form these groups was the number of years the RCS had been used, not the number of actually executed RCS-based assessments. These numbers seemed to be very rough estimates and as such unreliable. However, since it may be that even some intermediate level clinicians do not use the RCS much, it may be that the criterion of experience in years is still not totally satisfactory.

The decision to include participants from two countries was based on the fact of there not being enough experts in Finland, while at the same time the American experts could be more clearly defined as experts. While participants consisted of Finnish and American clinicians there is a possibility of intervening cultural factors, but the plethora of different educational RCS-programs, university programs and clinical programs does not make comparisons possible.

4.4.2 Design

There are four things in the nature of the design that limit generalizing from the results. First, the participants were not allowed to use any additional help, such as handbooks, and the assessment process had to be based solely on their mental contents and the given materials. That being the case, even if novices and intermediates made significantly more errors than experts, it does not necessarily mean that they do so in real life. Less experienced clinicians may use handbooks and supervision to correct errors in their thinking before writing reports. Hence, one cannot claim that in real life intermediate and novice level clinicians do a poorer job, although it remains equally possible that in real life the respective differences do exist. Second, since the study is based on comparisons, the results describe relationships between groups, not the relationship between a group and some other criterion. Thus, when a group has more errors than the expert group it does not mean that participants in the first group make too many errors.

Third, limitations in time given for evaluating materials may seem too short. However, after 25-35 minutes participants had nothing left to do and communicated that they were ready, which happened at all skill levels. Thus, in about 30 minutes they could do what was to be done based on their knowledge structures. It may be that in real life it takes more time to gather all the informa-

tion and write a report, than to think it through. Fourth, the main criterion for choosing the two specific clinical cases for this study was that those patients had been difficult to assess, which was thought to bring more variation between groups. However, in real life there are also easy cases, and therefore the chosen cases exaggerate the differences between groups.

The idea behind including the WAIS-R was that the design would have more ecological validity. Only the profile page was used to restrict the time used for examining it, but that was a clear pitfall which actually lessened the ecological validity. The results concerning the use of the WAIS-R profile do not have much information value, although at the same time this may also be because the profiles did not have much scatter.

Additionally, although it is common to talk about personality assessment, and we have the Journal of Personality Assessment, the Society for Personality Assessment, and so forth, some participants were puzzled about not having a more specific assessment question. This prompts the question about whether the task of assessing personality really mirrors actual clinical practice. Perhaps the title "personality assessment" is really just an abbreviation for many different kinds of more specific assessment tasks? Even if there is no definite answer to this question the task given in this study can still be used to study what kind of differences experience produces in the use of the RCS.

4.4.3 Problems in Gathering Data

There appeared a clear problem with the following instruction in the fourth phase of the data gathering method: "How would you now, in the light of the assessment you just made, understand or explain the patient's problems or symptoms". The rationale behind this question was to obtain information, first, on whether the participants would approach the materials by emphatically trying to understand the patient or by giving causal /scientific explanations, and second, to see whether they would be more concerned with patients' symptoms or some other problems. Many participants, however, did not understand this instruction, and indeed it had not been trialed in the pilots. Although this was a clear methodological shortcoming probably the only effect was that the results from this particular phase could not be used, as they were not.

Some technical problems were associated with the use of the equipment. Concerning protocols 10A and 10B, only the Description Phases could be utilized, because the volume of the voice was too low. With 16B the tape was empty. With 36A the tape recorder accidentally cut the tape, which was noticed about 1-5 minutes later. These shortcomings were noticed in the statistical analyses.

4.4.4 Categories

In this thesis the data has been examined by using a data-driven approach. This naturally prompts the question concerning the categories and their nature. The categories that were created during the analyses (item, inference, integration, error, content and structure of personality characteristics) seem to be credible in the sense that clinicians have to go through the materials, reason, integrate, and construct personality characteristics. Nevertheless, it is legitimate to ask whether they correctly reflect the nature of the data, especially in a data-driven exploratory study. This question takes three different forms. The first one is whether there are other possible categories? Indeed there are. With a rich body of data there are always a variety of different ways to approach it. The chosen categories bring with them conceptual commitments of the examiner, from which there is no escape. It is only natural that the researcher has a perspective to build analysis from, an awareness of some issues guiding the research questions, a school of thought to help sensitize the emergent concepts, or a degree of personal experience (O'Callaghan, 1996). The second form of the question is whether there are other relevant categories? Again, indeed there are. Examiners will always have their own personal paradigm and belief system, which to some extent dictate ontological and epistemological commitments (Goulding, 1998). If a researcher intends to study hospitals, he knows there will be doctors, nurses, and so on, and that their function is to facilitate proper diagnosis and cure.

The third form of the question is whether the categories correctly describe the chosen perspective. In this study there are two things that support a positive answer to this question. First, the classes were found and developed as a joint effort of four examiners. Second, the kappa coefficients support the idea that the phenomena are real. Nevertheless, methodologically, one can try to make the commitments transparent for the scientific community through an effort to make explicit what has been done and why. Whether the use of this kind of data-driven method is credible or not is a function of the extent to which the research process has been explicated. This way the scientific community will be conscious about why something is being hypothesized. An effort to reach this has been made in Chapter 2 and Appendices. Finally, our knowledge of the world does not depend on a single study but on the accumulation of results. When a study is accurately described the subsequent follow-up studies can match it as closely as possible.

Personality characteristics. A problem with the classes of personality characteristics is that they only classify characteristics on a general level and opposite characteristics (e.g., *depressed* and *not depressed*) belong to the same class. It was however rational to do this kind of analysis since there were only very few characteristics where such opposite views were given. In Case A there were such differences with five characteristics: 1) Eight participants thought there would be problems in thinking and four thought the opposite. 2) Eight participants though the patient would be anxious and one thought he was not. 3) Ten saw the patient as depressed and two not depressed. 4) Ten participants

saw the patient as not being able to cope or having small amount of resources, but three characterized him as resourceful and able to cope. 5) Five participants saw the patient as being in need of some heavier kind of treatment than psychotherapy, whereas one taught he could use psychotherapy. In Case B there was only one class with opposite views. Eleven participants saw that the patient has a low amount of resources and cannot cope well, but three thought she would have a lot of resources. These differences in Cases A and B, in total 66 personality characteristics, comprise about 10% of all characteristics. Thus, this particular problem is not fatal although in future analyses also a different kind of scheme can be developed.

Another thing about these classes is that they do not differentiate between personality characteristics that belong to the same class but are different, e.g., between *depressed* and *anxious* and *angry*, which all would belong to the class Experience (EXP). However, in this respect there was a problem that participants often did not characterize patients with single words, but used verbalizations that included more complex ways to depict the patient in some respect. This thesis resolved this by developing the classification system used here and also by analyzing the characteristics in terms of their structure.

The analyses of structures may be problematic in the sense that there was only one examiner, the implication being that these results can only be taken to indicate trends or hypotheses of structural differences. There were also some specific difficulties in coding structures. The structures that were quite easy to code were terms (i), negation (~), conjunction (&), modifier (M), and causality ($\rightarrow$). With predicates (P) the problematic aspect concerned verbs, since although some of these were simple to identify (such as *think, drink, believe*), with those that were more general natural language words (such as *do, go, is*) it was sometimes difficult to know whether they should be coded or not. Context (c) was often easy to identify (like *the fact that he was able to manage in the army,* or *she always ends up with these relationships were she gets abandoned*) since they repeated the anamnestic materials. With contexts however, there was a difficulty in saying whether a short phrase (e.g., *father*) should be taken as a context or not. Overall, the results concerning these structures are surely real in that there are structural differences in personality characteristics given by different groups, but one should be careful in thinking about what the differences more specifically are. However, this kind of analysis complements the results in an important way. It gives indication about expertise in this area, as well as how and why there may be very much divergence in real life assessment.

4.4.5 Representativeness and Generalization

The factors that are most problematic in terms of representativeness are the following. 1) The shortcoming of using only the profile page of the WAIS-R concluded in that the given materials do not represent cases where clinicians use a test battery. 2) Since the problem-solving task had to be executed without additional help, this may not correspond to the situation where especially less experienced clinicians use handbooks or supervision. 3) The definition of groups

(the number of years of using the RCS) may be problematic if there is a lot of variance in how much similarly experienced clinicians have executed the RCS-based personality assessments. 4) The given clinical cases were difficult and as such exaggerate the differences between groups. - Thus, one should be careful when generalizing the results in considering, e.g., the use of a test-battery or difference between differently experienced clinicians when facing simple or stereotypical clinical cases. Since there was only one coder in terms of the structures of personality characteristics, one should also give some consideration when generalizing them. However, the given problem-solving task is most probably adequate to tap on differently experienced participants' way of using the RCS in personality assessment, and the way participants execute the RCS-based personality assessment based on their mental contents.

4.5　Implications

4.5.1 Teaching

The empirical research on clinicians has lagged behind the research on personality tests. This has created a situation where the assessment process has become more like a theoretical entity or presupposition in the literature. Every clinician knows they have to think and process information during assessment, but nobody knows what this process really is like. While more research would make these processes more transparent, it would further facilitate teaching that is based on empirical research. To the best of my knowledge, there are no empirical studies on how to teach the RCS-based personality assessment. Thus, this section intends to present some provisional ideas about what, when, and how to teach.

What to teach? Concerning teaching the RCS, the literature contains little controversy as to what the cornerstones in teaching the RCS are. The three main areas are how to take a protocol, how to code it, and how to interpret it. Of these the one relevant here is interpretation, but to the best of my knowledge there are no empirical clinician-level studies in this area.

Generally, in a problem-solving area that necessitates thinking or information processing, the aim of teaching is that the students form knowledge structures based on which they can execute different mental operations. Here the essential operations are the use of Variables, making inferences, executing integration and forming structurally thorough personality characteristics. The knowledge structures are probably different in these four cases. In terms of using Variables, teaching should more forcefully acknowledge the importance of going through all Variables. While training usually aims to teach using all the information, participants fail to do this in practice. However, there is more to this since experts too often utilized only a portion of Variables, which was interpreted as them selecting what was relevant.

In terms of inferences, it is essential to guarantee that they do not include errors, which simply means to go back to the original definitions of Exner (2000, 2003). The results also show that all groups made more numerical than conceptual errors, implying it to be easier to handle some technical aspect of the RCS than to understand the meaning of Variables, which further brings forth the importance of understanding the conceptual contents of Variables. If novices and intermediates had understood Variables better, they probably had used more of them.

In terms of integration, while it naturally suffers when inferences are incorrect, personality assessment too suffers when there is not enough integration or when it is incorrect. The students should be taught that integration should not be executed if there are problems in understanding Variables. Only after the contents of Variables are understood completely, should the teaching pursue to help in giving explicit practices in integration.

Finally, based on the results it is in order to say something about how to teach clinicians to become experts in the RCS-based personality assessment. From an empirical standpoint it is proper, although otherwise quite trivial, to notice that to gain expertise means that one practices the areas pointed out in the results. A less trivial theoretical view is offered in the expertise research. Professional life usually comprises many different kinds of routine activities, which at some point become automatic. With these routine activities the mere addition in experience does not lead to expertise. If not experience, what makes an expert? Based on the expertise research, the answer is: context and amount of time devoted to deliberate practice. Deliberate practice means that "the most effective learning requires a well-defined task with an appropriate difficulty level for the particular individual, informative feedback, and opportunities for repetition and corrections of errors" (Ericsson, 1996, pp. 20-1). It is often done in total solitude and/or early in the morning in order to maximize concentration, and every day. For chess players practice does not mean playing chess, nor does it mean playing musical pieces for music performers. Similarly, in personality assessment with the RCS, expertise cannot be achieved by assessing people, but by using time and efforts to intentionally practicing the weak areas.

When to teach? This question can be divided into two different questions. The first one concerns the amount of time needed to be experienced enough to use the RCS. Basic training programs most probably are good enough in terms of administering and coding the RCS, one proof of which are the good results on studies concerning coding reliability (Exner, 2003). But interpreting and integrating are different matters. According to Kelley (1942), it takes at least 1-2 years for someone to be competent with the Rorschach. Klopfer (1943) saw that to learn interpretation takes about 1-5 years, while diagnostics cannot be taught. The training of the novices and intermediates in this study included 22 days of class work and lectures over the time span of 13-14 months, plus homework on administering, coding and interpretation with adequate feedback during the same time period. Although the results did not tell whether the errors were large enough for the practical personality assessment to suffer or not, it would

seem natural to work towards less erroneous assessment. Thus, 1 year is not enough for managing interpretation of Variables, and once this is learnt still more time should be devoted to integration.

Another question concerns the proper timing of teaching efforts. The similarity between intermediates and novices was an important finding, and in terms of teaching this can mean two things. It can mean that novices really have not acquired good enough skills and there is only little if any development after basic programs. In this case neither of the groups would be able to use the RCS well, and then the biggest task would be for training to facilitate better learning of the RCS. Another possible interpretation of this similarity is that basic training programs do give sufficient skills for everyday clinical practice, but it is very difficult to develop skills above that level. Both groups would perform well, but the development of expertise would be very laborious. In this case the focus should be on continuing education.

How to teach? Teaching can be pursued, e.g., by lectures, practices, reading and discussions. The results of this thesis do not as such give guidelines as to how to teach, but it is useful here to add some theoretical notions from expertise research to the discussion. The rationale behind this is that knowledge structures are an essential factor in errors, and expertise research gives one view in developing those. A way to approach this issue starts too with the concept of deliberate practice. Applied to personality assessment with the RCS, the essence of teaching would be to give appropriate individualized practices and feedback to facilitate the development of proper knowledge structures. The nature of a student's mental models must, of course, be known before setting practices with appropriate difficulty level, and thus before teaching there should be tests that show what he or she masters and what not. Following this a focus for practicing can be chosen. This new focus is again intentionally practiced to improve performance, which, in turn, is monitored, and so on. The importance of monitoring is in the fact that even if some specific area has been taught it does not mean it was learned. Whether that happened and to what extent it happened can only be known by testing it empirically. Thus, the use of deliberate practice together with individualized tests guarantees the development of adequate knowledge structures so that students really master what they should. It cannot be overemphasized that the errors in assessment are due to the problems in clinicians' knowledge structures, and therefore these should be the focus.

4.5.2 Research

The most general recommendation for future research stems from the exploratory nature of this study: to verify by further research that this study has grasped the essential phenomena and analyzed them appropriately. In addition to this there are many other specific areas where the research should be deepened or widened.

Use of test-items. This thesis did not detect differences in how differently experienced participants used the RCS-items, but this theme can be approached

further in a different manner by examining the nature of processes that the items comprise. Participants did not follow Exner's steps for interpretation (2003) but went through the Variables in some other order, and every now and then they went back to one of the earlier Variables. An example of a further research would be to trace the sequences of Variables with each participant and compare their nature. This would give important information about the nature of reasoning processes behind Variables, and also behind other RCS-materials.

Information contents. One evident lack of this thesis has been that information contents were not further analyzed. Inferences formed from Variables and during integration could be classified further based on their contents. Most probably this too would reveal something important about the thinking processes during the assessment process.

Information integration. One of the most important findings concerned the integration of information, especially integrating the RCS with anamnesis. This study only provided the general results in how groups differed, but it did not analyze further the nature of integration, and it would be important to provide these analyses. This would mean to study, e.g., whether there are different types of between-source integration, or what contents in the anamnesis integration episodes targeted? Further, there are questions concerning anamnesis in integration: How was anamnesis used? Do differently experienced clinicians select different pieces from it? Additionally, Exner (1993, 2003) has been stressing the importance of within-source integration, but the results here give only hints towards this being one phenomenon that differentiates experts and is thus demanding for other groups. This should also be further studied.

Different kinds of knowledge. Different types of knowledge and their interplay has been found an important area in the research on medical research. This should also be done in personality assessment. In practice this means to control the participant groups also in terms of clinical experience.

Other personality tests. To broaden the research in personality assessment, similar studies should be executed with other modern personality tests, e.g., MMPI-2 or PAI. If the number of tests in a study increased it would make it possible to execute ecologically more valid research, but this might be difficult in terms of the time per each participant and time devoted to analyses. Here only the RCS was used with history and the WAIS-R profile, which took about 45 minutes per patient. Using 2 cases and 30 subjects produced 610 pages of transcribed data, and there still were difficulties in having high enough base rates. Nevertheless, since the integration of personality tests has been a question of much interest lately, this kind of clinician-level study would in an important way add to this discussion.

Accuracy. The questions concerning the accuracy of personality assessment are highly relevant, but in this thesis only errors were evaluated, and no way was yet found to examine whether one personality characteristic was more accurate than the other. This however is a very important question since assessment naturally strives for accurate descriptions that are behind valuable decisions. Thus, another important area of future research would be to know

whether the novices and intermediates make too many errors, or qualitatively decisive errors, that endanger the process, or not. Or, what kinds of processes precede correct and incorrect characteristics?

Selectivity. In the expertise literature one important phenomenon with larger body of information has been selectivity, which means that experts are more skilled in selecting the relevant pieces of information (e.g., de Groot, 1965; Saariluoma, 1995). Concerning the RCS-based personality assessment it would be important to know whether and how experts select relevant information both from the anamnesis and from the RCS. Since personality assessment contains very large amounts of information, selectivity may become an extremely important activity.

Problem-solving tasks. The research in personality assessment can be broadened by studying the use of test-based information when different types of tasks are given. The task in this thesis was to assess personality, but also other types of questions are theoretically possible. However, there may be limitations to this. In the pilot studies of this thesis other tasks were tried with intermediate and novice level participants (personality dynamics, defenses, possible psychotic features, and/or treatment), but participants found the tasks too difficult. Also, in the fourth Phase of this study participants were asked to "explain or understand the patient's problems or symptoms?" Even though this question was thought to be reasonable, it didn't seem to make sense to the participants.

Skill development. Finally, it would be important to gain the knowledge as to why there were almost no differences between the novice and intermediate groups, that is, what causes arrested skill development. This piece of information would make it possible to tailor more effectively either the basic training process or the continuing education.

REFERENCES

Abrams, E. N. 1955. Prediction of intelligence from certain Rorschach factors. Journal of Clinical Psychology 11, 81-83.

Ackerman, S. J., Hilsenroth, M. J., Clemence, A. J., Weatherill, R. & Fowler, J. C. 2001. Convergent validity of Rorschach and TAT scales of object relations. Journal of Personality Assessment 77 (2), 295-306.

Affleck, D. C. & Mednick, S. A. 1959. The use of the Rorschach test in the prediction of the abrupt terminator in individual psychotherapy. Journal of Consulting Psychology 23 (2), 125-128.

Albert, S., Fox, H. M. & Kahn, M. W. 1980. Faking psychosis on the Rorschach: Can expert judges detect malingering? Journal of Personality Assessment 44 (2), 115-119.

Alpert, D. L. 1999. Integration of the MMPI and the Rorschach Ink Blot technique: A critical review. Dissertation Abstracts, 60 (2-B), 0816.

Ames, L. B. & Walker, R. N. 1964. Prediction of later reading ability from kindergarten Rorschach and IQ scores. Journal of Educational Psychology 55 (6), 309-313. Abstract.

Archer, R. P. & Krishnamurthy, R. 1997. MMPI-A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. Journal of Personality Assessment 69 (3), 517-533.

Archer, R. P. & Smith, S. R. 2008. Personality assessment. New York: Routledge.

Arffa, S. 1982. Predicting adolescent suicidal behavior and the order of Rorschach measurement. Journal of Personality Assessment 46 (6), 563-568. Abstract.

Argabrite, K. I. 1998. An exploratory study of the Rorschach Ego Impair Index as a predictor of adaptation to middle school. Dissertation Abstracts, 58 (10-B), 5693.

Armitage, S. G., Greenberg, P. D., Pearl, D., Berger, D. G. & Daston, P. G. 1955. Predicting intelligence from the Rorschach. Journal of Consulting Psychology 19 (5), 321-329.

Armitage, S. G. & Pearl, D. 1957. Unsuccessful differential diagnosis from the Rorschach. Journal of Consulting Psychology 21 (6), 479-484.

Auld, F. Jr. & Eron, L. D. 1953. The use of Rorschach scores to predict whether patients will continue psychotherapy. Journal of Consulting Psychology 17 (2), 104-109.

Barrows, H. S., Feightner, J. W., Neufeld, V. R. & Norman, G. R. 1978. Analysis of the clinical methods of medical students and physicians. Final Report, Ontario Department of Health Grants ODH-PR-273 & ODH-DM-226. Hamilton, Ontario: McMaster University. Abstract.

Barrows, H. S. & Tamblyn, R. M. 1980. Problem-based learning: An approach to medical education. New York: Springer.

Beck, S. J. 1930. Personality diagnosis by means of the Rorschach test. American Journal of Orthopsychiatry 1, 81-88. Abstract.

Belter, R. W. & Piotrowski, C. 2001. Current status of doctoral-level training in psychological testing. Journal of Clinical Psychology 57, 717–726.

Bernreuter, R. G. 1931. The theory and construction of the Personality Inventory. Palo Alto, CA: Consulting Psychologists Press.

Beutler, L. E. & Groth-Marnat, G. 2005. Integrative assessment of adult personality. (2nd edition) New York: The Guilford Press.

Bialick, I. & Hamlin, R. M. 1954. The clinician as judge: details of procedure in judging projective material. Journal of Consulting Psychology 18 (4), 239-242.

Biederman, L. & Cerbus, G. 1971. Changes in Rorschach teaching. Journal of Personality Assessment 35 (6), 524-526.

Blake, S. E., Humphrey, L. L. & Feldman, L. 1994. Self-delineation and marital interaction: The Rorschach predicts structural analysis of social behavior. Journal of Personality Assessment 63 (1), 148-166.

Boreham, N. C., Mawer, G. E. & Foster, R. W. 2000. Medical students' errors in pharmacotherapeutics. Medical Education 34 (3), 188-193.

Boshuizen, H. P. A. & Schmidt, H. G. 1992. Biomedical knowledge and clinical expertise. Cognitive Science 16, 153-184.

Breakwell, G. M., Hammond S. & Fife-Schaw, C. (Eds.) 2000. Research methods in psychology. (2nd edition). London, England: Sage.

Breuer, J. & Freud, S. 1895/1960. Studies in hysteria. Oxford, England: Beacon.

Bricklin, B. & Gottlieb, S. G. 1961. The prediction of some aspects of marital compatibility by means of the Rorschach test. Psychiatric Quarterly Supplement 35 (2), 281-303.

Buck, J. N. 1948. The H-T-P technique; a qualitative and quantitative method. Journal of Clinical Psychology 4, 317-396.

Bulbrook, M. E. 1932. An experimental study into the existence and nature of "insight". American Journal of Psychology 44, 409-453.

Burns, R. C. & Kaufman, S. H. 1970. Kinetic Family Drawings (K-F-D) research and application. New York: Brunner/Mazel.

Burns, R. C. & Kaufman, S. H. 1972. Actions, styles, and symbols in Kinetic Family Drawings (K-F-D). New York: Brunner/Mazel.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M. & Kaemmer, B. 1989. MMPI-2: Manual for administration and scoring. Minneapolis: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A. M., Ben-Porath, Y. S., et al. 1992. MMPI-A (Minnesota Multiphasic Personality Inventory-Adolescent): Manual for administration, scoring, and interpretation. Minneapolis: University of Minnesota Press.

Cantley, N. L. 2006. Rorschach predictors of outcome with a forensic conditional release population. Dissertation Abstracts, 67 (4-B), 2217.

Carlson, R. 1969. Rorschach prediction of success in clinical training: A second look. Journal of Consulting and Clinical Psychology 33 (6), 699-704.

Cartwright, R. D. 1958. Predicting response to client-centered therapy with the Rorschach PR Scale. Journal of Counseling Psychology 5 (1), 11-17.

Cates, J. A. 1999. The art of assessment in psychology: Ethics, expertise, and validity. Journal of Clinical Psychology 55 (5), 631-641.

Cattell, J. M. 1928. The founding of the association and of the Hopkins and Clark laboratories. Science 67, 546f.

Cattell, R. 1949. Handbook for the Sixteen Personality Factor Questionnaire. Champaign, IL: Institute for Personality and Ability Testing.

Cavanagh, S. 1997. Content analysis: Concepts, methods and applications. Nurse Researcher 4, 5-16.

Chambers, G. S. & Hamlin, R. M. 1957. The validity of judgments based on "blind" Rorschach records. Journal of Consulting Psychology 21 (2), 105-109.

Chase, W. G. & Simon, H. A. 1973a. The mind's eye in chess. In W. G. Chase (Ed.) Visual information processing. New York: Academic Press, 215-281.

Chase, W. G. & Simon, H. A. 1973b. Perception in chess. Cognitive Psychology 1, 33-81.

Chi, M. T. H. 2006. Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge University Press, 167-184.

Chi, M. T. H., Feltovich, P. & Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. Cognitive Science 5, 121-152.

Childs, R. A. & Eyde, L. D. 2002. Assessment training in clinical psychology doctoral programs: What should we teach? What do we teach? Journal of Personality Assessment 78, 130–144.

Chomsky, N. 1957/1985. Syntactic structures. The Hague: Mouton.

Cioffi, J. & Markham, R. 1997. Clinical decision-making by midwives: Managing case complexity. Journal of Advanced Nursing 25 (2), 265-272.

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

Claparéde, E. 1934. Genése de l'hypothése. Archives de Psychologie 24, 1-155. Abstract.

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, 18–47.

Coderre, S., Mandin, H., Harasym, P. H. & Fick, G. H. 2003. Diagnostic reasoning strategies and diagnostic success. Medical Education 37 (8), 695-703.

Cooper, J. G. 1955. The inspection Rorschach in the prediction of college success. Journal of Educational Research 49, 275-282.

Corcoran-Perry, S. A., Narayan, S. M. & Cochrane, S. 1999. Coronary care nurses' clinical decision making. Nursing & Health Sciences 1 (1), 49-61.

Costa, P. T. Jr. & McCrae, R. R. 1985. The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources.

Costa, P. T. Jr. & McCrae, R. R. 1992. Normal personality assessment in clinical practice: The NEO Personality Inventory. Psychological Assessment 4 (1), 5-13.

Crawford, E. F. 2004. Multi-method assessment of malingered depression and physical disability using the MMPI-2 and Rorschach inkblot method. Dissertation Abstracts, 65 (4-B), 2090.

Crowley, R. S., Naus, G. J. & Friedman, C. P. 2001. Development of visual diagnostic expertise in pathology. Proceedings/AMIA. Annual Symposium, 125-129.

Crowley, R. S., Naus, G. J., Stewart, J. 3rd. & Friedman, C. P. 2003. Development of visual diagnostic expertise in pathology - an information-processing study. Journal of the American Medical Informatics Association 10 (1), 39-51.

Cummings, S. T. 1954. The clinician as judge: Judgments of adjustment from Rorschach single-card performance. Journal of Consulting Psychology 18 (4), 243-247.

Dana, R. H., Cocking, R. R. & Dana, J. M. 1970. The effects of experience and training on accuracy and configural analysis. Journal of Clinical Psychology 26 (1), 28-32.

Dana, R. H. & Cunningham, K. M. 1983. Convergent validity of Rorschach and Thematic Apperception Test ego strength measures. Perceptual and Motor Skills 57 (3, Pt 2), 1101-1102.

Dao, T. K. 2008. Convergent and incremental validity of the MMPI-2 and Roschach on psychotic-related indices. Dissertation Abstracts, 68 (9-B), 6377.

Daston, P. G. & Sakheim, G. A. 1960. Prediction of successful suicide from the Rorschach test, using a sign approach. Journal of Projective Techniques 24, 355-361.

Dawes, R. M. 1994. House of cards: Psychology and psychotherapy built on myth. New York: Free Press.

Davids, A. & Talmadge, M. 1964. Utility of the Rorschach in predicting movement in psychiatric casework. Journal of Consulting Psychology 28 (4), 311-316.

Dey, I. 1993. Qualitative data analysis. User-friendly guide for social scientists. London: Routledge.

Dorr, D. 2008. Clinical integration of the MCMI-III and the Rorschach Comprehensive System. In T. Millon & C. Bloom (Eds.) The Millon inventories: A practitioner's guide to personalized clinical assessment. (2nd edition) New York, NY: Guilford Press, 135-156.

Duncker, K. 1926. A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). Pedagogical Seminary 33, 642-708.

Duncker, K. 1945. On problem solving. Psychological Monographs 58 (270), 1-113.

Durkin, H. E. 1937. Trial and error, gradual analysis, and sudden reorganization. An experimental study of problem solving. Archives of Psychology (Columbia University) 210, 85.

Elfhag, K., Rossner, S., Lindgren, T., Andersson, I. & Carlsson, A. M. 2004. Rorschach personality predictors of weight loss with behavior modification in obesity treatment. Journal of Personality Assessment 83 (3), 293-305.

Elstein, A. S., Shulman, L-S. & Sprafka, S. A. 1978. Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press

Endicott, N. E. & Endicott, J. 1964. Prediction of improvement in treated and untreated patients using the Rorschach prognostic rating scale. Journal of Consulting Psychology 28 (4), 342-348.

Epstein, S. A. 1991. Wage, labor, & guilds in medieval Europe. Chapel Hill, NC: North Carolina University Press.

Ericsson, K. A. 1996. The Acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.) The road to excellence. The acquisition of expert performance in the arts and sciences, sports and games. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1-50.

Ericsson, K. A. 2006a. An introduction to Cambridge Handbook of Expertise and Expert Performance: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 3-19.

Ericsson, K. A. 2006b. The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 683-703.

Ericsson, K. A. 2006c. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 223-241.

Ericsson, K. A. & Charness, N. 1997. Cognitive and developmental factors in expert performance. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.) Expertise in context: Human and machine. Menlo Park, CA, US: American Association for Artificial Intelligence; The MIT Press, 3-41.

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

Ericsson, K. A., Krampe, R. T. & Tesch-Römer, C. 1993. The role of deliberate practice in the acquisition of expert performance. Psychological Review 100 (3), 363-406.

Ericsson, K. A. & Simon, H. A. 1993. Protocol analysis. Verbal reports as data. Revised edition. Cambridge: The MIT Press.

Eva, K. W., Norman, G. R. & Brooks, L. R. 2002. Forward reasoning as a hallmark of expertise in medicine: logical, psychological, phenomeno-logical inconsistencies. Advances in Psychology Research 8, 25-40.

Exner, J. E. 1969. The Rorschach systems. New York: Grune & Stratton.

Exner, J. E. 1974. The Rorschach: A comprehensive system. Vol. 1. New York: Wiley.

Exner, J. E. 1978. The Rorschach: A comprehensive system. Vol. 2. Current research and advanced interpretation. New York: Wiley.

Exner, J. E. 1986. The Rorschach: A comprehensive system. Vol. 1. Basic foundations. (2nd edition) New York: Wiley.

Exner, J. E. 1991. The Rorschach: A comprehensive system. Vol. 2. Interpretation. (2nd edition) New York: Wiley.

Exner, J. E. 1993. The Rorschach: A comprehensive system. Vol. 1. Basic foundations. (3rd edition) New York: Wiley.

Exner, J. E. 2000. A primer for Rorschach interpretation. Asheville, North Carolina: Rorschach Workshops.

Exner, J. E. 2002. A new nonpatient sample for the Rorschach Comprehensive System: A progress report. Journal of Personality Assessment 78 (3), 391-404.

Exner, J. E. Jr. 2003. The Rorschach. A comprehensive system. Basic foundations and principles of interpretation. Hoboken, New Jersey: John Wiley & Sons.

Exner, J. E. Jr. & Erdberg, P. 2005. The Rorschach. A comprehensive system. Advanced Interpretation. Hoboken, New Jersey: John Wiley & Sons.

Exner, J. E. & Exner, D. E. 1972. How clinicians use the Rorschach. Journal of Personality Assessment 36, 403-408.

Farberow, N. L. 1974. Use of the Rorschach in predicting and understanding suicide. Journal of Personality Assessment 38 (5), 411-419.

Fassino, S., Amianto, F., Levi, M. & Rovera, G. G. 2003. Combining the Rorschach Test and the Temperament Character Inventory: A New Perspective on Personality Assessment. Psychopathology 36 (2), 84-91.

Faterson, H. F. & Klopfer, B. 1945. A survey of psychologists' opinions concerning the Rorschach method. Rorschach Research Exchange 9, 23-29.

Feltovich, P. J., Prietula, M. J. & Ericsson, K. A. 2006. Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. B. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 41-67.

Fernandez-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Hornke, L. F., Ter Laak, J. & Vizcarro, C., et al. 2001. Guidelines for the assessment process (GAP): A proposal for discussion. European Journal of Psychological Assessment 17, 187–200.

Filmer-Bennett, G. 1955. The Rorschach as a means of predicting treatment outcome. Journal of Consulting Psychology 19 (5), 331-334.

Finn, S. E. 1996. Assessment feedback integrating MMPI-2 and Rorschach findings. Journal of Personality Assessment 67 (3), 543-557.

Fisher, A. & Fonteyn, M. 1995. An exploration of an innovative methodological approach for examining nurses' heuristic use in clinical practice. Scholarly Inquiry for Nursing Practice 9 (3), 263-276.

Flavell, J. 1979. Metacognition and cognition monitoring: A new area of cognitive-developmental inquiry. American Psychologist 34, 906-911.

Fonteyn, M. E. 1997. Thinking in nursing practice. Studies in Health Technology & Informatics 46, 305-310.

Fonteyn, M. E. & Grobe, S. J. 1992. Expert nurses' clinical reasoning under uncertainty: Representation, structure, and process. Proceedings - the Annual Symposium on Computer Applications in Medical Care, 405-409.

Fonteyn, M. E. & Grobe, S. J. 1994. Expert system development in nursing: Implications for critical care nursing practice. Heart & Lung 23 (1), 80-87.

Fonteyn, M. & Fisher, A. 1995. Use of think aloud method to study nurses' reasoning and decision making in clinical practice settings. Journal of Neuroscience Nursing 27 (2), 124-128.

Forer, B. R., Farberow, N. L., Meyer, M. M. & Tolman, R. S. 1952. Consistency and agreement in the judgment of Rorschach signs. Journal of Projective Techniques 16, 346-351.

Fowler, L. P. 1997. Clinical reasoning strategies used during care planning. Clinical Nursing Research 6 (4), 349-361.

Frank, L. K. 1939. Projective methods for the study of personality. Transactions of the New York Academy of Sciences 1, 129-132.

Freud, S. 1899/1955. The interpretation of dreams. Oxford, England: Basic Books.

Gacono, C. B. & Meloy, J. R. 1994. The Rorschach assessment of aggressive and pychopathic personalities. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Ganellen, R. J. 1996a. Integrating the Rorschach and the MMPI-2 in personality assessment. Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Ganellen, R. J. 1996b. Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: A review. Journal of Personality Assessment 67 (2), 219-243.

Ganellen, R. J. 2001. Weighting evidence for the Rorschach's validity: A response to Wood et al. (1999). Journal of Personality Assessment 77 (1), 1-15.

Garb, H. N., Wood, J. M., Nezworski, M. T., Grove, W. M. & Stejkal, W. J. 2001. Towards the resolution of the Rorschach controversy. Psychological Assessment 13, 433-448.

Gardner, H. 1993. Creating minds: An anatomy of creativity seen through the lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham, and Gandhi. New York: Basic Books.

Gerstle, R. M., Geary, D. C., Himelstein, P. & Reller-Geary, L. 1988. Rorschach predictors of therapeutic outcome for inpatient treatment of children: A proactive study. Journal of Clinical Psychology 44 (2), 277-280.

Gibby, R. G., Stotsky, B. A., Hiler, E. W. & Miller, D. R. 1954. Validation of Rorschach criteria for predicting duration of therapy. Journal of Consulting Psychology 18 (3), 185-191.

Gibby, R. G., Stotsky, B. A., Miller, D. R. & Hiler, E. W. 1953. Prediction of duration of therapy from the Rorschach Test. Journal of Consulting Psychology 17 (5), 348-354.

Gilhooly, K. J. 1996. Thinking. Directed, undirected and creative. (3rd edition) New York: Academic Press.

Gilhooly, K. J., McGeorge, P., Hunter, J., Rawles, J. M., Kirby, I. K. & Green, C., et al. 1997. Biomedical knowledge in diagnostic thinking: The case of electrocardiogram (ECG) interpretation. European Journal of Cognitive Psychology 9 (2), 199-223.

Di Giulio, P. & Crow, R. 1997. Cognitive processes nurses and doctors use in the administration of PRN (at need) analgesic drugs. Scandinavian Journal of Caring Sciences 11 (1), 12-19.

Glaser, R. & Chi, M. T. H. 1988. Overview. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.) The nature of expertise. Hillsdale, NJ: Erlbaum, xv-xxviii.

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

Goodenough, F. L. 1926. Measurement of intelligence by drawings. Oxford, England: World Book.

Goulding, C. 1998. Grounded theory: The missing methodology on the interpretivist agenda. Qualitative Market Research: An International Journal 1 (1), 50-57.

Grant, M., Ives, V. & Ranzoni, J. 1952. Reliability and validity of judges' ratings of adjustment on the Rorschach. Psychological Monographs 66 (2, Whole No. 334).

Graves, P. L., Mead, L. A. & Pearson, T. A. 1986. The Rorschach Interaction Scale as a potential predictor of cancer. Psychosomatic Medicine 48 (8), 549-563.

Graves, P. L. & Thomas, C. B. 1981. Themes of interaction in medical students' Rorschach responses as predictors of midlife health or disease. Psychosomatic Medicine 43 (3), 215-225.

Grebstein, L. 1963. Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. Journal of Consulting Psychology 37, 127-132.

Greenwood, J., Sullivan, J., Spence, K. & McDonald, M. 2000. Nursing scripts and the organizational influences on critical thinking: Report of a study of neonatal nurses' clinical reasoning. Journal of Advanced Nursing 31 (5), 1106-1114.

Grobe, S. J., Drew, J. A. & Fonteyn, M. E. 1991. A descriptive analysis of experienced nurses' clinical reasoning during a planning task. Research in Nursing & Health 14 (4), 305-314.

de Groot, A. D. 1946/1965. Thought and choice in chess. The Hague: Mouton.

Groth-Marnat, G. 1999. Handbook of psychological assessment. (3rd edition) New York: Wiley.

Gruen, A. 1957. Rorschach: Some comments on predicting structured behavior from reactions to unstructured stimuli. Journal of Projective Techniques 21, 253-257.

Gunderson, E. K. & Kapfer, E. L. 1966. The predictive validity of clinical ratings for an extreme environment. British Journal of Psychiatry 112 (485), 405-412.

Hammond, K. R., Hursch, C. J. & Todd, F. J. 1964. Analyzing the components of clinical inference. Psychological Review 71, 438-456.

Handler, L. & Meyer, G. J. 1998. The importance of teaching and learning personality assessment. In L. Handler & M. J. Hilsenroth (Eds.) Teaching and learning personality assessment. Mahwah, New Jersey: Lawrence Earlbaum Associates, 3-30.

Hathaway, S. R. & McKinley, J. C. 1940. A multiphasic personality schedule (Minnesota): Pt. I. Construction of the schedule. Journal of Psychology 10, 249-254.

Hathaway, S. R. & McKinley, J. C. 1943. Minnesota Multiphasic Personality Inventory. Minneapolis: University of Minnesota.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R. & Brunell-Neuleib, S. 1999. A comparative meta-analysis of Rorschach and MMPI validity. Psychological Assessment 11, 278-296.

Hilsenroth, M. J. & Handler, L. 1995. A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. Journal of Personality Assessment 64 (2), 243-257.

Horn, J. & Masunaga, H. 2006. A merging theory of expertise and intelligence. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge University Press, 587-611.

Hughes, S. A., Deville, C., Chalhoub, M. & Romboletti, R. 1992. The Rorschach human anatomy response: Predicting sexual offending behavior in juveniles. Journal of Psychiatry & Law 20 (3), 313-333.

Hunt, E. 2006. Expertise, talent, and social encouragement. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. B. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 31-38.

Hunsley, J. & Bailey, J. M. 1999. The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. Psychological Assessment 11 (3), 266-277.

Jackson, D. N. 1974. Personality research form manual. Goshen, NY: Research Psychologists Press.

Janson, H. & Stattin, H. 2003. Prediction of adolescent and adult from childhood Rorschach ratings. Journal of Personality Assessment 81 (1), 51-63.

Jung, C. G. 1910. The association method. American Journal of Psychology 21, 219-269.

Kelley, D. M. 1942. Requirements for Rorschach training. Rorschach Research Exchange 6, 74-77.

Kendra, J. M. 1979. Predicting suicide using the Rorschach Inkblot Test. Journal of Personality Assessment 43 (5), 452-456.

Kestenbaum, J. M. & Lynch, D. 1978. Rorschach suicide predictors: A crossvalidational study. Journal of Clinical Psychology 34 (3), 754-758.

Klopfer, B. 1943. Instruction in the Rorschach method. Journal of Consulting Psychology 7 (2), 112-119.

Klopfer, B., Ainsworth, M. D., Klopfer, W. G. & Holt, R. R. 1954. Developments in the Rorschach technique. Volume 1. Technique and theory. New York: World Book.

Kohn, L. T., Corrigan, J. & Donaldson, M. S. 1999. To err is human: Building a safer heath system. Washington, DC: National Academy Press.

Kotkov, B. & Meadow, A. 1953. Rorschach criteria for predicting continuation in individual psychotherapy. Journal of Consulting Psychology 17 (1), 16-20.

Krause, E. A. 1996. Death of guilds: Professions, states and the advance of capitalism, 1930 to the present. New Haven, CT: Yale University Press.

Krishnamurthy, R., Archer, R. P. & House, J. J. 1996. The MMPI-A and Rorschach: A failure to establish convergent validity. Assessment 3 (2), 179-191.

Krishnamurthy, R., VandeCreek, L., Kaslow, N. J., Tazeau, Y. N., Miville, M. L. & Kerns, R., et al. 2004. Achieving competency in psychological assessment: Directions for education and training. Journal of Clinical Psychology 60 (7), 725–739.

Kushniruk, A. W. 2001. Analysis of complex decision-making processes in health care: cognitive approaches to health informatics. Journal of Biomedical Informatics 34 (5), 365-376.

Kushniruk, A., Patel, V. & Fleiszer, D. 1995. Analysis of medical decision making: A cognitive perspective on medical informatics. Proceedings - the Annual Symposium on Computer Applications in Medical Care, 193-197.

Kushniruk, A. W., Patel, V. L. & Marley, A. A. 1998. Small worlds and medical expertise: Implications for medical cognition and knowledge engineering. International Journal of Medical Informatics 49 (3), 255-271.

LaBarbera, J. D. & Cornsweet, C. 1985. Rorschach predictors of therapeutic outcome in a child psychiatric inpatient service. Journal of Personality Assessment 49 (2), 120-124.

Lamond, D., Crow, R. A. & Chase, J. 1996. Judgements and processes in care decisions in acute medical and surgical wards. Journal of Evaluation in Clinical Practice 2 (3), 211-216.

Larkin, J. H. 1983. The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.) Mental models. Mahwah, NJ: Lawrence Erlbaum Associates.

Lerner, P. M. 1998. Psychoanalytic perspectives on the Rorschach. Hillsdale, NJ: The Analytic Press.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D. & Wang, Y. 1988. Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.) The nature of expertise. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 311-342.

Lester, D., Kendra, J. M., Thisted, R. A. & Perdue, W. C. 1975. Prediction of homicide with the Rorschach. Journal of Clinical Psychology 31 (4), 752.

Levi, J. 1951. Rorschach patterns predicting success or failure in the re-habilitation of the physically handicapped. The Journal of Abnormal and Social Psychology 46 (2), 240-244.

Lisansky, E. S. 1956. The inter-examiner reliability of the Rorschach test. Journal of Projective Techniques 20, 310-317.

Loevinger, J. & Wessler, R. 1970. Measuring ego development: Vol. 1. Construction and use of a sentence completion test. San Francisco: Jossey-Bass.

Loevinger, J., Wessler, R. & Redmore, C. 1970. Measuring ego development: Vol. 2. Scoring manual for women and girls test. San Francisco: Jossey-Bass.

Long, E. & Miezitis, S. 1966. Prediction of aggressiveness in school children from clusters of signs on the Rorschach test. Ontario Journal of Educational Research 8 (3), 261-266.

Long, J. S. 1995. A comparative study between the Rorschach and MMPI-A in predicting symptoms and behaviors with inpatient adolescents. Dissertation Abstracts, 56 (6-B), 3452.

Lorenz, W., Schult, H. D & Rothmund, M. 1990. Chirurgische Entscheidungs-findung: Wechselspiel zwischen Erfahrung und klinischer Studie. Lange-becks Archiv fur Chirurgie - Supplement II - Verhandlungen der Deutschen Gesellschaft fur Chirurgie, 1283-1290. Abstract.

Lovitt, R. 1993. A strategy for integrating a normal MMPI-2 and dysfunctional Rorschach in a severely compromised patient. Journal of Personality Assessment 60 (1), 141-147.

Lubin, B., Larsen, R. M. & Matarazzo, J. D. 1984. Patterns of psychological test usage in the United States: 1935-1982. American Psychologist 39 (4), 451-454.

Machover, K. 1949. Personality projection in the drawing of the human figure. Oxford, England: C. C. Thomas.

Maslow, A. 1962. Toward a psychology of being. Princeton, N.J.: D. Van Nostrand Company.

McCandless, B. R. 1949. The Rorschach as a predictor of academic success. Journal of Applied Psychology 33 (1), 43-50.

McNamara, T. P. 1994. Knowledge representation. In R. J. Sternberg (Ed.) Thinking and problem solving. San Diego: Academic Press, 83-117.

Meehl, P. E. 1954. Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence. Minneapolis: University of Minnesota Press.

Meloy, J. R., Acklin, M. W., Gacono, C. B., Murray, J. F. & Peterson, C. A. 1997. Contemporary Rorschach interpretation. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Meyer, G. J. 1997. On the integration of personality assessment methods: The Rorschach and MMPI. Journal of Personality Assessment 68 (2), 297-330.

Meyer, G. J. 1999. Introduction to the special series on the utility of the Rorschach in clinical assessment. Psychological Assessment 11, 235-239.

Meyer, G. J. 2000. Incremental validity of the Rorschach Prognostic Rating scale over the MMPI Ego Strength Scale and IQ. Journal of Personality Assessment 74 (3), 356-370.

Meyer, G. J. & Archer, R. P. 2001. The hard science of Rorschach research: What do we know and where do we go? Psychological Assessment 13 (4), 486-502.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Kubiszyn, T. W. & Moreland, K., et al. 1998. Benefits and costs of psychological assessment in healthcare delivery: Report of the board of professional affairs psychological assessment work group, Part I. Washington, DC: American Psychological Association.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L. & Dies, R. R., et al. 2001. Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist 56, 128-165.

Meyer, G. J. & Handler, L. 1997. The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. Journal of Personality Assessment 69 (1), 1-38.

Miller, G. A., Galanter, E. & Pribram, K. H. 1960. Plans and the structure of behavior. New York, NY: Henry Holt.

Millon, T. 1977. MCMI manual. Minneapolis, MN: Interpretive Scoring Systems.

Mischel, W. 1968. Personality and assessment. New York: Wiley.

Morey, L. C. 1991. The Personality Assessment Inventory professional manual. Odessa, FL: Psychological Assessment Resources.

Morgan, C. D. & Murray, H. A. 1935. A method for investigating fantasies: The Thematic Apperception Test. Archives of Neurology and Psychiatry 34, 289-306.

Morris, W. W. 1943. Prognostic possibilities of the Rorschach method in metrazol therapy. American Journal of Psychiatry 100, 222-230.

Munroe, R. L. 1945. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. International Review of Applied Psychology 7, 104.

Murray, H. A. 1943. Thematic Apperception Test manual. Cambridge, MA: Harvard University Press.

Murstein, B. I. 1972. A thematic test and the Rorschach in predicting marital choice. Journal of Personality Assessment 36 (3), 213-217.

Neuringer, C. 1965. The Rorschach test as a research device for the identification, prediction and understanding of suicidal ideation and behavior. Journal of Projective Techniques & Personality Assessment 29 (1), 71-82.

Newell, A. & Simon, H. A. 1972. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall. Abstract.

Newton, R. L. 1954. The clinician as judge: Total Rorschachs and clinical case material. Journal of Consulting Psychology 18 (4), 248-250.

Nezworski, M. T. & Wood, J. M. 1995. Narcissicism in the comprehensive system for the Rorschach. Clinical Psychology: Science and Practice 2, 179-199.

Nickerson, R. S. 1994. The teaching of thinking and problem solving. In R. J. Sternberg (Ed.) Thinking and problem solving. San Diego: Academic Press, 409-449.

Nicolas, S. & Ferrand, L. 1999. Wundt's laboratory at Leipzig in 1891. History of Psychology 2 (3), 194-203.

Norman, G., Eva, K., Brooks, L. & Hamstra, S. 2006. Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. B. Hoffman (Eds.) The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press, 339-353.

Norman, G. R., Trott, A.L., Brooks, L. R. & Smith, E. K. M. 1994. Cognitive differences in clinical reasoning related to postgraduate training. Teaching and Learning in Medicine 6, 114-120.

O'Callaghan, J. 1996. Grounded theory: a potential methodology. Counselling Psychology Review 11 (1), 23-28.

Offredy, M. 2002. Decision-making in primary care: outcomes from a study using patient scenarios. Journal of Advanced Nursing 40 (5), 532-541.

Patel, V. L. & Arocha, J. F. 2001. The nature of constraints on collaborative decision-making in health care settings. In E. Salas & G. A. Klein (Eds.) Linking expertise and naturalistic decision making. Mahwah, NJ: Erlbaum, 383-405.

Patel, V. L., Arocha, J. F., Diermeier, M., How, J. & Mottur-Pilson, C. 2001. Cognitive psychological studies of representation and use of clinical practice guidelines. International Journal of Medical Informatics 63 (3), 147-167.

Patel, V. L., Arocha, J. F. & Zhang, J. 2005. Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.) The Cambridge handbook of thinking and reasoning. Cambridge: Cambridge University Press, 727-750.

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

Patel, V. L. & Groen, G. J. 1986b. Non-monotonicity in novice-intermediate-expert comparisons. In Proceedings of the 27th Annual Meeting of the Psychonomic Society, New Orleans, LA.

Patel, V. J. & Groen, G. J. 1991a. Developmental accounts of the transition from medical students to doctor: Some problems and suggestions. Medical Education 25, 527-535.

Patel, V. L. & Groen, G. J. 1991b. The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.) Toward a general theory of expertise: Prospects and limits. New York: Cambridge University Press, 93-125.

Perry, W., McDougall, A. & Viglione, D. 1995. A five-year follow-up on the temporal stability of the Ego Impairment Index. Journal of Personality Assessment 64 (1), 112-118.

Perry, W. & Potterat, E. G. 1997. Beyond personality assessment: The use of the Rorschach as a neuropsychological instrument in patients with amnestic disorders. In J. R. Meloy, M. W. Acklin, C.B. Gacono & J. F. Murray (Eds.) Contemporary Rorschach interpretation. Mahwah, NJ, US: Lawrence Erlbaum Associates, 557-575.

Perry, W., Potterat, E., Auslander, L. & Kaplan, E., 1996. A neuropsychological approach to the Rorschach in patients with dementia of the Alzheimer type. Assessment 3 (3), 351-363.

Perry, W. & Viglione, D. J. 1991. The Ego Impairment Index as a predictor of outcome in melancholic depressed patients treated with tricyclic antidepressants. Journal of Personality Assessment 56 (3), 487-501.

Perry, W., Viglione, D. & Braff, D. 1992. The Ego Impairment Index and schizophrenia: A validation study. Journal of Personality Assessment 59 (1), 165-175.

Peterson, D. R. 1968. The clinical study of social behavior. East Norwalk, CT, US: Appleton-Century-Crofts.

Piotrowski, C. & Zalewski, C. 1993. Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. Journal of Personality Assessment 61, 394–405.

Potkay, C. R. & Ward, E. F. 1972. Clinical judgment under varied informational conditions: Rorschach, personal data, and best guess. Journal of Consulting and Clinical Psychology 39 (3), 513.

Powers, W. T. & Hamlin, R. M. 1957. The validity, bases, and process of clinical judgment, using a limited amount of projective test data. Journal of Projective Techniques 21, 286-293.

Pressey, L. W. 1918. Sex differences shown by 2544 school children on a group scale of intelligence, with special reference to variability. Journal of Applied Psychology II, 323-340.

Pressey, S. L. & Pressey, L. W. 1919. "Cross-out" tests with suggestions as to a group scale of the emotions; studies from the psychological laboratory of Indiana University. Journal of Applied Psychology 3 (2), 138-150.

Psykologi [Psychologist] 2008. Psykologisten testien käyttö Suomessa laajaa. [The use of psychological tests is wide in Finland.] 8, 18-19.

Rader, G. E. 1957. The prediction of overt aggressive verbal behavior from Rorschach content. Journal of Projective Techniques 21, 294-306.

Redden, M. & Wotton, K. 2001. Clinical decision making by nurses when faced with third-space fluid shift. How well do they fare? Gastroenterology Nursing 24 (4), 182-191.

Reitman, W. R. 1965. Cognition and thought. New York: John Wiley.

Richman, H. B., Gobet, F., Staszewski, J. J. & Simon, H. A. 1996. Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.) The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games. Mahwah: Erlbaum, 167-187.

Ritzler, B. & Alter, B. 1986. Rorschach teaching in APA-approved clinical graduate programs: Ten years later. Journal of Personality Assessment 50 (1), 44-49.

Ritzler, B. A. & del Gaudio, A. C. 1976. A survey of Rorschach teaching in APA-approved clinical graduate programs. Journal of Personality Assessment 40 (5), 451-453.

Roberts, L. K. 1954. The failure of some Rorschach indices to predict the outcome of psychotherapy. Journal of Consulting Psychology 18 (2), 96-98.

Rogers, C. R. 1961. On becoming a person. Oxford, England: Houghton Mifflin.

Rogers, L. S. & Hammond, K. R. 1953. Prediction of the results of therapy by means of the Rorschach test. Journal of Consulting Psychology 17 (1), 8-15.

Rogers, L. S., Knauss, J. & Hammond, K. R. 1951. Predicting continuation in therapy by means of the Rorschach Test. Journal of Consulting Psychology 15 (5), 368-371.

Rorschach, H. 1921/1942. Psychodiagnostics: A diagnostic test based on perception. Bern, Switzerland: Hans Huber.

Rosenthal, R., Hiller, J. B., Bornstein, R. F., Berry, D. T. R. & Brunell-Neuleib, S. 2001. Meta-analytic methods, the Rorschach, and the MMPI. Psychological Assessment 13, 449-451.

Rotter, J. B. & Rafferty, J. E. 1950. Manual for the Rotter Incomplete Sentences Blank: College form. New York: Psychological Corporation.

Rotter, J. B., Rafferty, J. E. & Schachtitz, E. 1949. Validation of the Rotter Incomplete Sentences Blank for college screening. Journal of Consulting Psychology 13, 348-356.

Rotter, J. B. & Willerman, B. 1947. The Incomplete Sentences Test as a method of studying personality. Journal of Consulting Psychology 11, 42-48.

Russell, B. 1905/1956. On denoting. In B. Russell, Logic and knowledge. London: Allen and Onwin, 41-56.

Saariluoma, P. 1995. Chess players' thinking. A cognitive psychological approach. New York: Routledge.

Schafer, R. 1954. Psychoanalytic interpretation in Rorschach testing: theory and application. Oxford, England: Grune & Stratton.

Shaffer, T. W., Erdberg, P. & Haroian, J. 1999. Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. Journal of Personality Assessment 73, 305-316.

Shalit, B. 1965. Effects of environmental stimulation on the M, FM, and m responses in the Rorschach. Journal of Projective Techniques & Personality Assessment 29 (2), 228-231.

Silberg, J. L. & Armstrong, J. G. 1992. The Rorschach test for predicting suicide among depressed adolescent inpatients. Journal of Personality Assessment 59 (2), 290-303.

Simpson, S. A. & Gilhooly, K. J. 1997. Diagnostic thinking processes: Evidence from a constructive interaction study of electrocardiogram (ECG) interpretation. Applied Cognitive Psychology 11, 543-554.

Sloan, W. 1948. Prediction of extramural adjustment of mental defectives by use of the Rorschach test. Journal of Consulting Psychology 12 (5), 303-309.

Smith, S. R. & Archer, R. P. 2008. Introducing personality assessment. In R. P. Archer & S. R. Smith (Eds.) Personality Assessment. New York: Routledge, 1-36.

Society for Personality Assessment 2005. The status of the Rorschach in clinical and forensic practice: An official statement by the Board of Trustees of the Society for Personality Assessment. Journal of Personality Assessment 85, 219-237

Society for Personality Assessment 2006. Standards for education and training in psychological assessment: Position of the Society for Personality Assessment: An official statement of the Board of Trustees of the Society for Personality Assessment. Journal of Personality Assessment 87 (3), 355-357.

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J. & Hayes, A. 1996. Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.) The road to excellence. The acquisition of expert performance in the arts and sciences, sports and games. Mahwah, New Jersey: Lawrence Erlbaum Associates, 81-106.

Stedman, J. M., Hatch, J. P. & Schoenfeld, L. S. 2001. The current status of psychological assessment training in graduate and professional schools. Journal of Personality Assessment 77, 398–407.

Sterling, S. B. 1995. Adolescent Rorschach predictors of adult problem drinking. Dissertation Abstracts, 55 (7-B), 3028.

Stokes, A. F., Kemper, K. & Kite, K. 1997. Aeronautical decision making, cue recognition, and expertise under time pressure. In C. E. Zsambok & G. Klein (Eds.), Naturalistic decision making: Research and applications. Mahwah, NJ: Erlbaum, 183-196.

Stokes, J. M., Pogge, D. L., Powell-Lunder, J., Ward, A. W., Bilginer, L. & DeLuca, V. A. 2003. The Rorschach Ego Impairment Index: Prediction of treatment outcome in a child psychiatric population. Journal of Personality Assessment 81 (1), 11-19

Sulfaro, M. M. 2005. Predicting adolescent suicidality with a new Rorschach Index: A cross-validation study. Dissertation Abstracts, 65 (10-B), 5424.

Suppes, P. 1957. Introduction to logic. New York: Van Nostrand.

Taulbee, E. S. 1955. The use of the Rorschach test in evaluating the intellectual levels of functioning in schizophrenics. Journal of Projective Techniques 19, 163-169.

Thompson, G. M. 1948. College grades and the group Rorschach. Journal of Applied Psychology 32 (4), 398-407.

Thompson, G. M. 1951. College grades and the group Rorschach: a follow-up study. Journal of Genetic Psychology 78, 39-46.

Tuber, S. B. 1983. Children's Rorschach scores as predictors of later adjustment. Journal of Consulting and Clinical Psychology 51 (3), 379-385.

Turner, D. R. 1966. Predictive efficiency as a function of amount of information and level of professional experience. Journal of Projective Techniques and Personality Assessment 30, 4-11.

Vandenbert, S. G., Rosenzweig, N., Moore, K. R. & Dukay, A. F. Diagnostic agreements among psychiatrists and "blind" Rorschach raters of the education of an interdisciplinary research team. Psychological Reports 15 (2), 211-224.

Viglione, D. J. 1999. A review of recent research addressing the utility of the Rorschach. Psychological Assessment 11, 251-265.

Viglione, D. J. & Hilsenroth, M. J. 2001. The Rorschach: Facts, fictions, and future. Psychological Assessment 13, 452-471.

Viglione, D. J. & Taylor, N. 2003. Empirical support for interrater reliability of Rorschach comprehensive system coding. Journal of Clinical Psychology 59 (1), 111-121.

Watkins, C. E. 1991. What have surveys taught us about the teaching and practice of psychological assessment? Journal of Personality Assessment 56, 426–437.

Watson, J. B. 1920. Is thinking merely the action of language mechanisms? British Journal of Psychology 11, 87-104.

Weilbaecher, A. N. 2002. MMPI-2 and Rorschach indices of schizophrenia: An evaluation of convergent and incremental validity. Dissertation Abstracts, 62 (8-B), 3818.

Weiner, I. B. 1993. Clinical considerations in the conjoint use of the Rorschach and the MMPI. Journal of Personality Assessment 60 (1), 148-152.

Weiner, I. B. 1998. Teaching the Rorschach Comprehensive System. In L. Handler & M. J. Hilsenroth (Eds.) Teaching and learning personality assessment. Mahwah, New Jersey: Lawrence Earlbaum Associates, 215-233.

Weiner, I. B. & Greene, R. L. 2008. Handbook of personality assessment. Hoboken, New Jersey: John Wiley & Sons.

Winslow, C. N. & Rapersand, I. 1964. Postdiction of the outcome of somatic therapy from the Rorschach records of schizophrenic patients. Journal of Consulting Psychology 28 (3), 243-247.

Witmer, L. 1907. Clinical psychology. American Psychologist 1, 1-9.

Wood, J. M., Nezworski, M. T., Garb, H. N. & Lilienfeld, S. O. 2001. The misperception of psychopathology: Problems with the norms of the

comprehensive system for the Rorschach. Clinical Psychology: Science and Practice 8, 350-373.

Wood, J. M., Nezworski, M. T. & Stejkal, W. J. 1996. The comprehensive system for the Rorschach: A critical examination. Psychological Science 7, 3-10.

Wood, J. M., Nezworski, M. T., Stejkal, W. J., Garven, S. & West, S. G. 1999. Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner (1996), and Ganellen (1996). Assessment 6, 115-129.

Woodworth, R. S. 1919. Examination of emotional fitness for war. Psychological Bulletin 15, 59-60.

Woodworth, R. S. 1920. Personal Data Sheet. Chicago: Stoelting.

Young, I. L. 1973. Rorschach animal content as a predictor of aggressive behavior. Psychology: A Journal of Human Behavior 10 (3), 50-54.

TIIVISTELMÄ

Eksperttiys ja Rorschach Comprehensive System'in käyttö persoonallisuuden arvioinnissa

Terveydenhuollon alueella ihmisten terveydentilan tutkiminen sisältää kaksi keskeistä elementtiä: tutkimisessa käytetyt välineet, ja noita välineitä käyttävät kliinikot. Kliinisessä persoonallisuuden arvioinnissa tätä erottelua vastaavat persoonallisuustestien käyttö ja persoonallisuuden arviointi. Testaaminen tarkoittaa sitä prosessia, jossa persoonallisuustestien avulla saadaan informaatiota ihmisen persoonallisuudesta. Persoonallisuuden arviointi taas viittaa prosessiin, jossa testaamisen tuottamaa informaatiota käsitellään päämääränä ihmisen ymmärtäminen ja ongelmien selvittäminen. Jotta kliininen persoonallisuuden arviointi olisi luotettavaa, näiden molempien puolien täytyy olla empiirisesti tutkittuja. Kliinisessä psykologiassa tilanne on kuitenkin huono sikäli, että vaikka monista persoonallisuustesteistä on olemassa huomattavan paljon empiiristä tutkimusta, persoonallisuuden arvioinnin prosessista ei ole tieteellistä tutkimusta. Tällaisen tutkimuksen puuttuminen tieteellisestä psykologiasta on ollut keskeinen motivaatio tähän väitöskirjatutkimukseen.

Lääketieteessä tilanne on parempi. Lääketieteellisiä instrumentteja on tutkittu huomattavan paljon, mutta viimeisten 30 vuoden aikana on tutkittu myös lääkäreitä heidän selvittäessään potilaiden ongelmia noiden instrumenttien avulla. Tuon tutkimuksen mukaan keskeiset ongelmat lääketieteellisessä diagnostiikassa liittyvät virheisiin ajattelussa. Tilastojen mukaan lääketieteellinen virhe onkin kahdeksanneksi yleisin kuolinsyy USA:ssa – yleisempi kuin kuolemat liikenteessä tai kuolemat rintasyöpään.

Myös virheet kliinisessä persoonallisuuden arvioinnissa vaikuttavat ihmisten elämään tuottamalla vääriä päätöksiä ja tarpeetonta kärsimystä. Se tosiasia, että testien tuottaman informaation prosessoinnista kliinisessä persoonallisuuden arvioinnissa ei käytännöllisesti katsoen ole lainkaan tieteellistä tietoa, on huolestuttavaa, ottaen huomioon sen, että persoonallisuustestejä käytetään erittäin paljon. Esimerkiksi Suomessa 84% psykologeista käyttää testejä, ja he testaavat noin 232 500 henkilöä vuodessa. Koska ei ole tutkimusta siitä, miten kliiniset psykologit käsittelevät persoonallisuustestien tuottamaa informaatiota, ei todellisuudessa ole olemassa tietoa siitä, mitä todellisuudessa tapahtuu tällä kliinisen psykologian alueella. Jos lääketieteessä tehdyt tutkimukset näistä prosesseista pätevät persoonallisuuden arviointiin, niin tuolloin persoonallisuuden arviointi on niin voimakkaasti riippuvainen kliinikoiden kyvystä käsitellä tietoa, että ilman tieteellistä tietoa joidenkin persoonallisuustestien erinomaisetkaan psykometriset ominaisuudet eivät voi pelastaa arviointia ajatusvirheiltä.

Tämän tutkimuksen tarkoituksena on ollut tutkia sitä, miten kliinisten psykologien taito vaikuttaa heidän kykyynsä käyttää Rorschach Comprehensive System'iä (RCS) persoonallisuuden arvioinnissa. Tältä alueelta ei ole olemassa aikaisempia tutkimuksia, joten tämä tutkimus on luonteeltaan eksploratiivi-

nen pyrkien tuottamaan tietoa, joka voi toimia myöhemmän tutkimuksen perustana. RCS on valittu tutkimukseen mm. siksi, että se on hyvin laajalti käytetty ja sen käyttö on vaikeaa.

Eksperttiystutkimus tarjoaa tehokkaan tavan lähestyä tätä aihepiiriä. Tässä metodologisessa lähestymistavassa eritasoisesti kokeneita koehenkilöitä verrataan heidän suorittaessaan ongelmanratkaisun tehtävää. Tämä antaa tietoa siitä millaista on välineen hyvä käyttö ja mitkä kohdat siinä ovat kaikkein eniten ongelmallisia. Siten tämän väitöskirjan pääkysymys on se, kuinka eritasoisesti kokeneet kliiniset psykologit käyttävät RCS:ää suorittaessaan persoonallisuuden arviointia sen tuottaman informaation avulla. Aihe on erittäin relevantti ajatellen mielenterveyspalvelujen laatua ja tuloksellisuutta.

Koehenkilöt muodostivat kolme taidoiltaan erilaista ryhmää (ekspertit, keskitasoiset, ja noviisit), joista jokainen sisälsi kymmenen koehenkilöä. Ryhmät määriteltiin sen mukaan kuinka monta vuotta he olivat käyttäneet RCS:ää. Eksperttien kokemus oli 9-30 vuotta, keskiryhmän 3-9 vuotta, ja kaikkien noviisien yksi vuosi. Koehenkilöiden tehtävänsä oli arvioida kaksi potilasta käyttäen potilaan taustatietoja, WAIS-R –profiilia ja RCS:n tuottamaa materiaalia. Arvioinnit suoritettiin ääneen ajatellen (ns. protokolla-analyysin menetelmä) ja litteroituja protokollia (yhteensä 610 sivua) käytettiin datana.

Koehenkilöt saivat ensin ääneen ajatellen käydä vapaasti läpi materiaaleja noin puoli tuntia, minkä jälkeen heitä pyydettiin kertomaan heidän näkemyksensä potilaan persoonallisuudesta. Tutkija kirjoitti ylös luetellut persoonallisuuden piirteet, minkä jälkeen hän kysyi koehenkilöltä mihin kohtiin materiaaleissa kukin piirre perustui.

Dataa analysoitiin sekä kvalitatiivisesti että kvantitatiivisesti. Keskeisten analyysiluokkien muodostaminen tapahtui kvalitatiivisesti. Neljä tutkijaa kävi läpi 3-4 vuoden aikana dataa päätyen seuraaviin luokkiin: RCS:n tuottamasta materiaalista käytetyt informaatioyksiköt; informaatiosta tehdyt päätelmät; informaation integrointi; prosessin aikana tehdyt virheet; sekä muodostettujen persoonallisuuden luonnehdintojen sisältö ja rakenne. Tämän kvalitatiivisen osuuden jälkeen koehenkilöryhmiä vertailtiin kvantitatiivisesti tilastoanalyysien avulla.

Tuloksissa tuli esille kaksi keskeistä asiaa. Ensiksi, ekspertit erosivat muista useissa suhteissa. He osasivat virheettömämmin käyttää RCS:n muuttujia, kykenivät paremmin yhdistämään RCS:n tuottamaa informaatiota muihin informaatiolähteisiin, käyttivät RCS:n protokollan vastauksia eri tavoin, integroivat muuttujia toisiin muuttujiin virheettömämmin, tekivät päätelmiä eri tavoin ja tuottivat kompleksisempia persoonallisuuden luonnehdintoja. Toinen keskeinen tulos oli se, että keskiryhmän ja noviisien välillä ei ollut mitään tilastollisesti merkitseviä eroja, mikä herättää kysymyksen taitojen kehittymisen luonteesta tällä alueella.

Tulosten yleistämisessä koskemaan käytännön kliinistä persoonallisuuden arviointia kannattaa olla varovainen. Tutkimuksessa koehenkilöt eivät saaneet käyttää mitään apuvälineitä, vaan arvioinnin tuli perustua pelkästään annettuihin materiaaleihin ja koehenkilön tietorakenteisiin. Työssään kliinikot voivat

kuitenkin käyttää apunaan kirjoja ja työnohjausta, jolloin erot saattavat tasoittua. Lisäksi, koska tutkimus perustuu ryhmien välisille vertailuille, tulosten perusteella ei voida väittää, että tutkimuksessa heikommin pärjänneet ryhmät tekisivät arkielämässä huonompaa työtä. Edustavuuden kannalta on oltava varovainen sikäli, että tutkimukseen valittiin vaativia potilastapauksia, kun käytännön työ usein sisältää myös vähemmän vaativia arviointitehtäviä.

Tutkimus on kuitenkin adekvaatti tapa lähestyä kliinikoiden kykyä käsitellä testien tuottamaa informaatiota omien tietorakenteidensa tai mentaalisten malliensa avulla. Se herättää ajatuksia persoonallisuuden arvioinnin opettamisesta ja siitä, kuinka voitaisiin saavuttaa sellainen taito, joka takaa potilaiden luotettavan arvioinnin.

Avainsanat: Eksperttiys, Rorschach Comprehensive System, persoonallisuuden arviointi, persoonallisuustesti, ääneen ajattelu, protokolla-analyysi.

APPENDIX 1 Case A

<u>Anamnesis</u>

This is a 24-year old male, graduated from high school some time ago, who comes to treatment with a referral from the health care center because of depression. There are no previous psychiatric treatments. He tells that his problems began some years ago, when his parents (who had divorced at that time) got into financial troubles and he gave up his savings to help them. After high school he had been working and saving money to buy himself an apartment, and bought a BMW, but when his mother told him about the parents financial difficulties and that their rents would be distrained, he took all the money from his savings account, paid some of his parents debts and worked in two shifts to help them. At the same time his friend totally crashed his car, so he lost both his savings and his car. His financial situation started to be bad and he had to give up his rented apartment and move to his mother. He says that at this time he had a nervous breakdown and his problems began. To relief his bad feeling he started to drink more than usual. The drinking increased so much that he could not always take care of things and was caught drunken driving. However, he could stop drinking too much by going to a short treatment in a clinic for alcoholics and getting appropriate psychiatric medication.

He has now been on a medical leave for more than a year, since his "nervous breakdown". He mostly spends time home at his mothers' place, without seeing people, doing nothing special or sometimes watching video movies whole day. His mother does all the homework. He tells he is so ashamed of his situation that he cannot go out. He has no telephone and he doesn't open his mail. Mother opens his mail and gives him all the most important letters. He tells he gets anxious very easily, and for example, if he gets bad news through mail, an invitation to trial or a larger bill, he gets anxiety attacks. In the bus he may get into panic and forget his things in there. He also tells he suffers from chronic insomnia. Related to anxiety and insomnia he has constantly used both sleeping pills and tranquillizers more than his doctor ordered.

He is the older of two children. His sister, a couple of years older than he, works in an office and has a boyfriend. His parents divorced 4 years ago. Mother is from Russia and unemployed for the moment. Father is an alcoholic and violent, and their relationship is bad. Apparently his father also suffers from severe mental problems, but has not been in any treatment, and has long been without apartment since the divorce. Occasionally the father visits the patient and ex-wife, probably to get to stay inside overnight. Then he often goes into one of the rooms and sits there in the dark and without saying a word for a couple of days until disappears somewhere. The patient experiences these situations very distressing. He doesn't want to talk to his father, and when his father rarely says something, he criticizes his son of being mentally ill or incapable of taking care of things, among other things.

The patient went normally through comprehensive school and high school, where his achievements were average. In the beginning of high school

he felt things were going pretty well. He had some hobbies, like taekwondo, badminton, fishing and weight-lifting. Because of the poor financial situation of their family, in the end of high school he had to work to get some money, as a guard and as a taxi driver. There were major financial disagreements in the family. According to the parents they couldn't afford to buy him school books, but then also he himself did not want to use his earnings in buying books, and he says he managed by listening in the classroom. The high school time was difficult, and sometimes he worked around the clock, sometimes even without eating at all. After high school he had some short employments for example in a warehouse, as a guard and driving a taxi. He did his compulsory military service in 2001, and he says he was then in very bad nutritional condition and gained 20 kilos during the army. Life was like vacation in the army when he had some distance to his father, because they continually had strong fights. After the army he worked 3 months in a warehouse and finally in an office, as part of a rehabilitating work activity of the employment agency, but that was interrupted because of psychological difficulties. He thought he would try to get to a technical school, but he couldn't go to the entrance exam because of his symptoms.

When he is at the psychologists' office his hands are all wet because he is so tensed, but in spite of this he tells things in an organized way, openly and in good co-operation. He doesn't bring up any suicidal thoughts and there is no impression of any psychotic features. He tells he is constantly in a low mood and totally helpless about his future. He feels everyday routine things are overwhelmingly difficult and normally he just stays at home.

WAIS-R RECORD FORM

WECHSLER ADULT
INTELLIGENCE SCALE—
REVISED

BRITISH ADAPTATION

OCCUPATION _____

COUNTRY OF BIRTH _____

PLACE OF TESTING _____

OTHER INFORMATION _____

NAME _____

ADDRESS _____

SEX _____ AGE _____ MARITAL STATUS _____

EDUCATION _____

FIRST LANGUAGE _____

TESTED BY _____

TABLE OF SCALED SCORE EQUIVALENTS*

Scaled Score	VERBAL TESTS						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	8-10	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.

Date block

	Year	Month	Day
Date Tested	—	—	—
Date of Birth	—	—	—
Age	—	—	—

SUMMARY

	Raw Score	Scaled Score
VERBAL TESTS		
Information		9
Digit Span		8
Vocabulary		7
Arithmetic		11
Comprehension		9
Similarities		12
Verbal Score		56
PERFORMANCE TESTS		
Picture Completion		7
Picture Arrangement		11
Block Design		7
Object Assembly		8
Digit Symbol		7
Performance Score		40

	Sum of Scaled Scores	IQ
VERBAL	56	95
PERFORMANCE	40	85
FULL SCALE	96	90

Structural Summary

Location Features

Zf	=	7
ZSum	=	22.5
ZEst	=	20.5
W	=	9
(Wv	=	3)
D	=	6
W+D	=	15
Dd	=	4
S	=	3

DQ

			(FQ-)
+	=	2	(0)
o	=	12	(1)
v/+	=	0	(0)
v	=	5	(1)

Form Quality

		FQx	MQual	W+D
+	=	0	0	0
o	=	8	1	7
u	=	9	0	6
-	=	2	1	2
none	=	0	0	0

Determinants

Blends	Single	
C.M	M	= 1
m.CF	FM	= 0
m.CF	m	= 0
	FC	= 2
	CF	= 0
	C	= 0
	Cn	= 0
	FC'	= 1
	C'F	= 0
	C'	= 0
	FT	= 0
	TF	= 0
	T	= 0
	FV	= 0
	VF	= 0
	V	= 0
	FY	= 0
	YF	= 0
	Y	= 0
	Fr	= 0
	rF	= 0
	FD	= 0
	F	= 12
	(2)	= 4

Contents

H	= 2
(H)	= 0
Hd	= 1
(Hd)	= 0
Hx	= 0
A	= 5
(A)	= 0
Ad	= 1
(Ad)	= 0
An	= 0
Art	= 2
Ay	= 0
Bl	= 0
Bt	= 5
Cg	= 0
Cl	= 0
Ex	= 0
Fd	= 0
Fi	= 1
Ge	= 1
Hh	= 0
Ls	= 0
Na	= 0
Sc	= 3
Sx	= 0
Xy	= 0
Idio	= 4

S-Constellation

☐	FV+VF+V+FD > 2
☐	Col-Shd Blends > 0
☑	Ego < .31 or > .44
☐	MOR > 3
☐	Zd > ±3.5
☐	es > EA
☑	CF + C > FC
☑	X+% < .70
☐	S > 3
☐	P < 3 or > 8
☐	Pure H < 2
☐	R < 17
3	Total

Special Scores

		Lvl-1	Lvl-2
DV	=	3 x1	0 x2
INC	=	0 x2	0 x4
DR	=	0 x3	0 x6
FAB	=	1 x4	1 x7
ALOG	=	0 x5	
CON	=	0 x7	

Raw Sum6 = 5
Wgtd Sum6 = 14

AB	= 0	GHR	= 1
AG	= 0	PHR	= 2
COP	= 0	MOR	= 2
CP	= 0	PER	= 0
		PSV	= 0

RATIOS, PERCENTAGES, AND DERIVATIONS

R = 19		L = 1.71

EB	= 2 : 4.5	EA	= 6.5	EBPer	= N/A
eb	= 2 : 1	es	= 3	D	= +1
		Adj es	= 2	Adj D	= +1

FM	= 0	SumC'	= 1	SumT	= 0
m	= 2	SumV	= 0	SumY	= 0

AFFECT

FC:CF+C	= 2 : 3
Pure C	= 1
SumC' : WSumC	= 1 : 4.5
Afr	= 0.46
S	= 3
Blends:R	= 3 : 19
CP	= 0

INTERPERSONAL

COP = 0		AG = 0
GHR:PHR		= 1 : 2
a:p		= 3 : 1
Food		= 0
SumT		= 0
Human Content		= 3
Pure H		= 2
PER		= 0
Isolation Index		= 0.32

IDEATION

a:p	= 3 : 1	Sum6	= 5
Ma:Mp	= 1 : 1	Lvl-2	= 1
2AB+(Art+Ay)	= 2	WSum6	= 14
MOR	= 2	M-	= 1
		M none	= 0

MEDIATION

XA%	= 0.89
WDA%	= 0.87
X-%	= 0.11
S-	= 0
P	= 3
X+%	= 0.42
Xu%	= 0.47

PROCESSING

Zf	= 7
W:D:Dd	= 9:6:4
W : M	= 9 : 2
Zd	= +2.0
PSV	= 0
DQ+	= 2
DQv	= 5

SELF-PERCEPTION

3r+(2)/R	= 0.21
Fr+rF	= 0
SumV	= 0
FD	= 0
An+Xy	= 0
MOR	= 2
H:(H)+Hd+(Hd)	= 2 : 1

PTI = 0	☐ DEPI = 4	☐ CDI = 2	☐ S-CON = 3	☐ HVI = No	☐ OBS = No

Sequence of Scores

Card	Resp. No	Location and DQ	Loc. No.	Determinant(s) and Form Quality	(2)	Content(s)	Pop	ZScore	Special Scores
I	1	Wo	1	Fo		A	P	1.0	
	2	WSv	1	Fo		Bt			MOR
II	3	Do	6	Fu		A			
	4	DSv	5	Fo		Id			MOR
III	5	D+	1	Mpo	2	H,Sc	P	3.0	GHR
	6	Ddo	99	Fu		A			
IV	7	Ddo	99	Fu		A			
	8	Do	6	Fo		Hd			PHR
V	9	Wv	1	Fu		Bt			
	10	Wo	1	Fo		A	P	1.0	DV
VI	11	Wo	1	Fu		Bt		2.5	DV
VII	12	Wv	1	Fu	2	Ge			DV
	13	WS+	1	FC'o		Id,Sc		4.0	FAB
VIII	14	Do	6	FCu		Bt			
IX	15	Ddo	99	FCu		Bt,Art			
	16	Wo	1	Fu	2	Art,Ad		5.5	
X	17	Wo	1	C.Ma-		H,Id		5.5	FAB2. PHR
	18	Dv	6	ma.CF-	2	Id			
	19	Ddo	99	ma.CFo		Fi.Sc			

Summary of Approach

I : W.WS	VI : W
II : D.DS	VII : W.WS
III : D.Dd	VIII : D
IV : Dd.D	IX : Dd.W
V : W.W	X : W.D.Dd

CONSTELLATIONS TABLE

S-Constellation (Suicide Potential)

☐ Positive if 8 or more conditions are true:
NOTE: Applicable only for subjects over 14 years old.

- ☐ FV+VF+V+FD [0] > 2
- ☐ Col-Shd Blends [0] > 0
- ☑ Ego [0.21] < .31 *or* > .44
- ☐ MOR [2] > 3
- ☐ Zd [2.0] > ±3.5
- ☐ es [3] > EA [6.5]
- ☑ CF + C [3] > FC [2]
- ☑ X+% [0.42] < .70
- ☐ S [3] > 3
- ☐ P [3] < 3 or > 8
- ☐ Pure H [2] < 2
- ☐ R [19] < 17

3   Total

PTI (Perceptual-Thinking Index)

- ☐ (XA% [0.89] < 0.70) *and* (WDA% [0.87] < 0.75)
- ☐ X-% [0.11] > 0.29
- ☐ (Sum Level 2 Special Scores [1] > 2)
  and (FAB2 [1] > 0)
- ☐ ((R [19] < 17) *and* (WSum6 [14] > 12)) *or*
  ((R [19] > 16) *and* (WSum6 [14] > 17))
- ☐ (M- [1] > 1) *or* (X-% [0.11] > 0.40)

0   Total

DEPI (Depression Index)

☐ Positive if 5 or more conditions are true:

- ☐ (FV + VF + V [0] > 0) *or* (FD [0] > 2)
- ☑ (Col-Shd Blends [0] > 0) *or* (S [3] > 2)
- ☑ (3r + (2)/R [0.21] > 0.44 *and* Fr + rF [0] = 0)
  or (3r + (2)/R [0.21] < 0.33)
- ☑ (Afr [0.46] < 0.46) *or* (Blends [3] < 4)
- ☐ (SumShading [1] > FM + m [2])
  or (SumC' [1] > 2)
- ☐ (MOR [2] > 2) *or* (2xAB + Art + Ay [2] > 3)
- ☑ (COP [0] < 2)
  or ([Bt+2xCl+Ge+Ls+2xNa]/R [0.32] > 0.24)

4   Total

CDI (Coping Deficit Index)

☐ Positive if 4 or more conditions are true:

- ☐ (EA [6.5] < 6) *or* (AdjD [1] < 0)
- ☑ (COP [0] < 2) *and* (AG [0] < 2)
- ☐ (Weighted Sum C [4.5] < 2.5)
  or (Afr [0.46] < 0.46)
- ☐ (Passive [1] > Active + 1 [4])
  or (Pure H [2] < 2)
- ☑ (Sum T [0] > 1)
  or (Isolate/R [0.32] > 0.24)
  or (Food [0] > 0)

2   Total

HVI (Hypervigilance Index)

☐ Positive if condition 1 is true and at least 4 of the others are true:

- ☑ (1) FT + TF + T [0] = 0
- ☐ (2) Zf [7] > 12
- ☐ (3) Zd [2.0] > +3.5
- ☐ (4) S [3] > 3
- ☐ (5) H + (H) + Hd + (Hd) [3] > 6
- ☐ (6) (H) + (A) + (Hd) + (Ad) [0] > 3
- ☑ (7) H + A : Hd + Ad [7:2] < 4 : 1
- ☐ (8) Cg [0] > 3

OBS (Obsessive Style Index)

- ☑ (1) Dd [4] > 3
- ☐ (2) Zf [7] > 12
- ☐ (3) Zd [2.0] > +3.0
- ☐ (4) Populars [3] > 7
- ☐ (5) FQ+ [0] > 1

☐ Positive if one or more is true:

- ☐ Conditions 1 to 5 are all true
- ☐ Two or more of 1 to 4 are true *and* FQ+ [0] > 3
- ☐ 3 or more of 1 to 5 are true
  and X+% [0.42] > 0.89
- ☐ FQ+ [0] > 3 *and* X+% [0.42] > 0.89

NOTE: '' indicates a cutoff that has been adjusted for age norms.*

SCZI (Schizophrenia Index)

☐ Positive if 4 or more conditions are true:

☑ ((X+% [0.42] < 0.61) *and* (S-% [0.00] < 0.41))
  or (X+% [0.42] < 0.50)

☐ X-% [0.11] > 0.29

☐ (FQ- [2] ≥ FQu [9])
  or (FQ- [2] > FQo [8] + FQ+ [0])

☐ (Sum Level 2 Special Scores [1] > 1)
  and (FAB2 [1] > 0)

☐ (Raw Sum of 6 Special Scores [5] > 6) *or*
  (Weighted Sum of 6 Special Scores [14] > 17)

☐ (M- [1] > 1)
  or (X-% [0.11] > 0.40)

| Total

Interpretive Hypotheses

The first step in considering the possible interpretive significance of Rorschach findings consists of determining whether a protocol is sufficiently long (more than 13 responses) and complete (no card rejections) to be useful; that is, to provide reliable data and support valid inferences. Additionally in this initial process, concerns should be raised if elevations in the S-CON suggest suicide potential. Interpretive cautions may be raised if highly unusual features of the data suggest efforts to simulate serious psychological disturbance.

This record contains a sufficient number of responses to provide reliable information and to support valid interpretations.

Constellations

PTI = 0 No	☐ DEPI = 4	☐ CDI = 2	☐ S-CON = 3	☐ HVI = No	☐ OBS = No

Interpretive Search Strategy

Positive Key Variable(s)	Interpretive Search Strategy
1. Lambda > 0.99	Processing > Mediation > Ideation > Controls > Affect > Self-Perception > Interpersonal Perception

<u>Rorschach Comprehensive System protocol</u>

I   1.   A bug, butterfly

          E: Repeats subject's Response
          S: Here is something like a bit of a feeler and these look like wings, here's the middle body

    2.   Could be a leaf that is a little bit ripped

          E: Repeats subject's Response
          S: It just came, a couple of holes in a leaf, it's worn and rubbed
          E: Worn and rubbed?
          S: The edges are ripped, there are holes in the middle
          E: Ripped?
          S: Yes, the edges are like that

II   3.   Well, nothing (subject gives the ink blot back, but the rejection is not accepted after which he continues): looks a little bit like a butterfly

          E: Repeats subject's Response
          S: These look a little bit like wings and this looks a little bit like a head
          E: Could you show where you saw it?
          S: This black area

    4.   Something that is ripped here in the middle

          E: Repeats subject's Response
          S: Here in the middle
          E: Ripped?
          S: I don't know, there's some hole or then that butterfly just has a white splash over there

III   5.   2 persons sitting together, there's something in the middle

          E: Repeats subject's Response
          S: Those look something like persons, here's something in the middle, they are a little bit like in that show "10 Thousand Dollar Questions", there are computers or something
          E: Sitting together?
          S: Yes they kind of sit facing each other

    6. (v) Some bug

          E: Repeats subject's Response
          S: Here are the legs, teeth, eyes
          E: Where do you see it?
          S: This upper part

IV   7. (v) Some bug

          E: Repeats subject's Response
          S: Again like wings on both sides and these are like feelers
          E: Where do you see it?
          S: It's divided from here

    8.   These two could be like legs

          E: Repeats subject's Response
          S: These two
          E: I'm not sure I understand what made it look like that?

S: This part, it's kind of oblong

V   9.   A leaf

E: Repeats subject's Response
S: Looks a little bit like a leaf
E: What made it look like a leaf?
S: These parts here

10.   A butterfly

E: Repeats subject's Response
S: These are like feelers, here are the wings, and these are like leglers

VI   11.   Could be a flower

E: Repeats subject's Response
S: Here's the stem and there are leaves below and then here's a bigger lump
E: Lump?
S: Or is it a flower … some plant anyway

VII   12. (v) Reminds me of 2 continents

E: Repeats subject's Response
S: Looks like it, here's one part of world and here's another
E: I'm not sure I understand what made it look like it?
S: I couldn't figure out anything else, they don't look like specific continents, but it could be some kind of mainland somewhere

13. (v) A mushroom with ink been thrown on it

E: Repeats subject's Response
S: It came from a mushroom, there has been a mushroom and ink has been thrown on it and it has left this kind of trace, so that this is ink
E: I'm not yet sure if I see what made it look like that?
S: It could be paint, that dark color

VIII   14. (v) Some kind of flower

E: Repeats subject's Response
S: This looks a little bit like a flower, these leafs, it's a little bit variegated
E: Variegated?
S: The upper part is variegated and lower part is greenish
E: I'm not yet sure if I understand what made it look like a flower?
S: It's this way, greenish leafs and this colored looks like a flower

IX   15. (v) Primarily it reminds me of a rose, all I can figure out is a rose that is daubed by a five year old

E: Repeats subject's Response
S: These look like petals of a rose, over here like there are a couple of green petals in roses
E: I'm not yet sure I understand what made it a rose?
S: It just came to my mind, they are kind of round, look a little bit sharp, and a little bit wide

16. (v) Could be also a coat-of-arms if there are 2 heads of bears over here

E: Repeats subject's Response
S: This is like an ear of a bear, there are two alike, could also be a stamp
E: What made it coat-of-arms?
S: A seal comes to my mind because there are usually 2 heads of bears or eagles in a seal, aren't there

X   17. (v) Over here could be a flying person

           E: Repeats subject's Response
           S: Here are the wings, there is a bit of a leg, here's the head, it's quite a sparkling picture
           E: Sparkling?
           E: A lot of colors have been thrown in there, a kind of psychedelic figure

       18.   Maybe some kind of an electrical reaction in the middle

           E: Repeats subject's Response
           S: This looks like 2 electrical discharges
           E: I still don't see what made it look like electrical discharges?
           S: It's kind of a blue and complicated thing

       19.   Maybe some fireworks

           E. Repeats subject's Response
           S: Well, this could be like some parts of a rocket banging
           E: Could you show me where it is?
           S: These parts that are the most colorful, not these dark ones, that could be a part of the rocket

ORDER FOR REVIEWING VARIABLES WITHIN EACH CLUSTER (rev 8/00)

CONTROL & STRESS TOLERANCE
Step 1 - Adjusted D Score and CDI
Step 2 - EA
Step 3 - EB and Lambda
Step 4 - es and Adj es
Step 5 - eb

SITUATION RELATED STRESS
Step 1 - D Score in relation to es and Adj es
Step 2 - Difference between D and Adj D Scores
Step 3 - m & Y
Step 4 - T, V+3r+(2)/R in relation to History
Step 5 - D Score (if appropriate Pure C, M¬, M no form)
Step 6 - Blends
Step 7 - Color Shading Blends

AFFECTIVE FEATURES
Step 1 - DEPI & CDI
Step 2 - EB & Lambda
Step 3 - EBPer
Step 4 - Right Side eb & variables related to it
Step 5 - SumC':WSumC
Step 6 - Affective Ratio
Step 7 - Intellectualization Index
Step 8 - Color Projection
Step 9 - FC:CF+C
Step 10- Pure C
Step 11- Space responses
Step 12- Blends (Lambda & EB)
Step 13- m & Y blends
Step 14- Blend complexity
Step 15- Color shading blends
Step 16- Shading blends

INFORMATION PROCESSING
Prerequisites (L,EB,OBS,HVI)
Step 1 - Zf
Step 2 - W:D:Dd
Step 3 - Location Sequencing
Step 4 - W:M
Step 5 - Zd
Step 6 - PSV
Step 7 - DQ
Step 8 - DQ Sequencing

MEDIATION
Prerequisites (R,OBS,L)
Step 1 - XA% & WDA%
Step 2 - FQnone
Step 3 - X-%, FQ- frequency, S- frequency
Step 3a- Homogeneity issues
Step 3b- Minus distortion levels
Step 4 - Populars
Step 5 - FQ+ frequency
Step 6 - X+% & Xu%

IDEATION
Step 1 - EB & Lambda
Step 2 - EBPer
Step 3 - a:p
Step 4 - HVI, OBS, MOR
Step 5 - Left side eb
Step 6 - Ma:Mp
Step 7 - Intellectualization Index
Step 8 - Sum6 & WSum6
Step 9 - Quality 6 Spec Scores
Step 10- M Form Quality
Step 11- Quality of M responses

SELF PERCEPTION
Step 1 - OBS & HVI
Step 2 - Reflections
Step 3 - Egocentricity Index
Step 4 - FD and Vista (in relation to History)
Step 5 - An+Xy
Step 6 - Sum MOR
Step 7 - H:(H)+Hd+(Hd) & Review codings
          for Human Content responses
Step 8 - Search for projections in:
Step 8a- Minus responses
Step 8b- MOR responses
Step 8c- M & human content responses
Step 8d- FM & m responses
Step 8e- Embellishments in other responses

INTERPERSONAL PERCEPTION
Step 1 - CDI
Step 2 - HVI
Step 3 - a:p Ratio
Step 4 - Food responses
Step 5 - Sum T
Step 6 - Sum Human Contents & Sum Pure H
Step 7 - GHR:PHR
Step 8 - COP & AG frequencies & codings
Step 9 - PER
Step 10- Isolation Index
Step 11- Contents of M & FM responses with pairs

APPENDIX 2 Case B

<u>Anamnesis</u>

This is a 26-year old young woman who lives alone, and who came to psychiatric treatment for the first time and because of depression. In addition to depression she brought up strong feelings of hopelessness and emptiness, and suicidal thoughts. The events leading to the beginning of treatment proceeded as follows: In the end of 2001 she started to feel worse and there started to be difficulties in sleeping and troubles at work. She was a licensed practical nurse and worked as an assistant taking care of children in day care. During the day she was tired, got easily irritated to screaming children and her relationship with another worker started to come to a head. As a result she resigned. At the same time she had problems in private life. She fell in love with a woman with husband and two children. She visited that family often, acting like a friend of the family and sometimes taking care of the couple's kids. The husband knew nothing about the relationship between his wife and the patient. The patient did not feel satisfied with how their relationship proceeded and she started to suspect that the woman would not leave her family. When she came to treatment she felt her life was like a dead end. In appearance she was clearly overweight, looking expressionless and unassuming, and always so meager in her talking, that for those treating her it was difficult to get a picture of what was going on and what caused her symptoms. She had difficulties in naming causes for her suicidal thoughts and anxiety. She brought up feelings of emptiness and difficulties to find meaning and goals in her life. She thought life was so heavy that dying would be an easier option and then she wouldn't have to think of all the difficult things.

After her resignation she has mostly spent time by being alone at home, which has easily aroused intense anxiety and loneliness. Sometimes she has met her only good friend, a young male homosexual, and her sister, but sometimes she has felt very conflicted about meeting her sister. Also the contacts with her mother have aroused very conflicted feelings and her mother, as being worried about her daughter's health, has been calling her daily. Once when her mother participated at a meeting in the hospital an arrangement was reached, that she would call her daughter only every other day, but soon she returned to daily calls. Sometime later, when the patient was at the hospital ward, she met another patient, again a married woman with two children. They met each other for a while and when this woman broke up, the patient started again to have more intense suicidal thoughts.

Also a couple of months after she had started her treatment and she was abroad with her mother, she started to have suicidal thoughts and she made a concrete plan of how to kill herself with medicine. She felt that there are too much all kinds of difficulties in life and she couldn't go on any more. At this time the treatment in the hospital was started. It took three months, and also later she was in the hospital many times for suicidal thoughts and intense anxiety. During the three months in hospital there were many unsuccessful dis-

charges. According to her symptoms, she had medication for anxiety, depression and problems in sleeping. When she wasn't at the ward she had therapy in an outpatient clinic on a weekly basis. When the problems didn't seem to ease she received electric shocks after which her suicidal thoughts markedly lessened.

She is the younger of two daughters. Her mother works in a factory and her father is retired because of medical reasons. The father has some problems with drinking. Her older sister, who lives together with another woman, lives nearby and is healthy. Early in the treatment the patient told that her sister behaved violently towards her during all their childhood in a manner that was not only some skirmishing between sisters. Much later she brought up, that her mother would have sexually abused her during many years when she was between 9 and 12. She also suspects that something like that may have happened earlier, but this is something she cannot remember. She has not been discussing about these matters either with her sister or with her mother, and in the treatment she mentions these things only shortly, but doesn't want to talk more about them. She's had two hobbies, cross-stitching and drawing, in both of which she is specialized in making pictures of teddy bears.

WAIS-R RECORD FORM

WECHSLER ADULT
INTELLIGENCE SCALE
REVISED

BRITISH ADAPTATION

NAME _____

ADDRESS _____

SEX _____ AGE _____ MARITAL STATUS _____

OCCUPATION _____ EDUCATION _____

COUNTRY OF BIRTH _____ FIRST LANGUAGE _____

PLACE OF TESTING _____ TESTED BY _____

OTHER INFORMATION *The subject is peaceful, concentrates well, and is observant. She understands the instructions and does her best, but does not overtry.*

TABLE OF SCALED SCORE EQUIVALENTS*

Scaled Score	VERBAL TESTS						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.

Date block

	Year	Month	Day
Date Tested			
Date of Birth			
Age			

SUMMARY

	Raw Score	Scaled Score
VERBAL TESTS		
Information		11
Digit Span		10
Vocabulary		11
Arithmetic		10
Comprehension		12
Similarities		6
Verbal Score		60
PERFORMANCE TESTS		
Picture Completion		9
Picture Arrangement		10
Block Design		12
Object Assembly		12
Digit Symbol		11
Performance Score		54

	Sum of Scaled Scores	IQ
VERBAL	60	99
PERFORMANCE	54	102
FULL SCALE	114	101

Structural Summary

Location Features

Zf	=	15
ZSum	=	62.5
ZEst	=	49.0
W	=	10
(Wv	=	0)
D	=	3
W+D	=	13
Dd	=	4
S	=	8

DQ

			(FQ-)
+	=	7	(1)
o	=	9	(7)
v/+	=	1	(1)
v	=	0	(0)

Form Quality

		FQx	MQual	W+D
+	=	0	0	0
o	=	2	2	2
u	=	6	4	6
-	=	9	6	5
none	=	0	0	0

Determinants

Blends	Single	
M.FC'	M	= 9
M.FD	FM	= 0
M.FC'	m	= 0
m.CF.FD.TF	FC	= 0
FM.CF	CF	= 0
FD.m.C	C	= 0
	Cn	= 0
	FC'	= 0
	C'F	= 0
	C'	= 0
	FT	= 0
	TF	= 0
	T	= 0
	FV	= 0
	VF	= 0
	V	= 0
	FY	= 0
	YF	= 0
	Y	= 0
	Fr	= 0
	rF	= 0
	FD	= 0
	F	= 2
	(2)	= 5

Contents

H	=	3
(H)	=	1
Hd	=	5
(Hd)	=	1
Hx	=	1
A	=	3
(A)	=	1
Ad	=	0
(Ad)	=	0
An	=	0
Art	=	1
Ay	=	0
Bl	=	1
Bt	=	0
Cg	=	1
Cl	=	0
Ex	=	0
Fd	=	0
Fi	=	2
Ge	=	0
Hh	=	1
Ls	=	0
Na	=	2
Sc	=	2
Sx	=	0
Xy	=	0
Idio	=	1

S-Constellation

☑	FV+VF+V+FD > 2
☑	Col-Shd Blends> 0
☑	Ego < .31 or > .44
☐	MOR > 3
☑	Zd > ±3.5
☐	es > EA
☑	CF+C > FC
☑	X+% < .70
☑	S > 3
☑	P < 3 or > 8
☐	Pure H < 2
☐	R < 17
8	Total

Special Scores

			Lvl-1	Lvl-2
DV	=		0 x1	0 x2
INC	=		2 x2	0 x4
DR	=		0 x3	0 x6
FAB	=		0 x4	1 x7
ALOG	=		0 x5	
CON	=		0 x7	

Raw Sum6 = 3
Wgtd Sum6 = 11

AB	= 0		GHR	= 3
AG	= 3		PHR	= 11
COP	= 0		MOR	= 1
CP	= 0		PER	= 0
			PSV	= 0

RATIOS, PERCENTAGES, AND DERIVATIONS

R = 17	L = 0.13

EB	= 12 : 3.5	EA = 15.5	EBPer = 3.4
eb	= 3 : 3	es = 6	D = +3
		Adj es = 5	Adj D = +4

FM	= 1	SumC' = 2	SumT = 1
m	= 2	SumV = 0	SumY = 0

AFFECT

FC:CF+C	= 0 : 3
Pure C	= 1
SumC' : WSumC	= 2 : 3.5
Afr	= 0.42
S	= 8
Blends:R	= 6 : 17
CP	= 0

INTERPERSONAL

COP = 0		AG	= 3
GHR:PHR			= 3 : 1
a:p			= 7 : 8
Food			= 0
SumT			= 1
Human Content			= 10
Pure H			= 3
PER			= 0
Isolation Index			= 0.24

IDEATION

a:p	= 7 : 8	Sum6	= 3
Ma:Mp	= 4 : 8	Lvl-2	= 1
2AB+(Art+Ay)	= 1	WSum6	= 11
MOR	= 1	M-	= 6
		M none	= 0

MEDIATION

XA%	= 0.47
WDA%	= 0.62
X-%	= 0.53
S-	= 7
P	= 2
X+%	= 0.12
Xu%	= 0.35

PROCESSING

Zf	= 15
W:D:Dd	= 10:3:4
W : M	= 10 : 12
Zd	= +13.5
PSV	= 0
DQ+	= 7
DQv	= 0

SELF-PERCEPTION

3r+(2)/R	= 0.29
Fr+rF	= 0
SumV	= 0
FD	= 3
An+Xy	= 0
MOR	= 1
H:(H)+Hd+(Hd)	= 3 : 7

PTI = 3	☑ DEPI = 5	☐ CDI = 1	☑ S-CON = 8	☐ HVI = No	☐ OBS = No

Sequence of Scores

Card	Resp. No	Location and DQ	Loc. No.	Determinant(s) and Form Quality	(2)	Content(s)	Pop	ZScore	Special Scores
I	1	W+	1	Mpu	2	A		4.0	INC, PHR
	2	WSo	1	Mp.FC'-		(Hd)		3.5	PHR
II	3	WSo	1	Ma-		Hd		4.5	MOR, PHR
	4	DdSo	99	Mp-		Hd		4.5	PHR
III	5	D+	1	Mao	2	H,Hh	P	3.0	AG, GHR
	6	Do	1	Mpu		A			INC, PHR
IV	7	W+	1	Ma.FDo		H,Sc	P	4.0	AG, GHR
V	8	DdSo	99	Mp-		Hd		4.0	PHR
	9	Do	10	Ma-	2	Hd			PHR
VI	10	W+	1	Fu		Na,Art		2.5	
	11	W+	1	Mpu	2	H		2.5	GHR
VII	12	DdSo	99	F-		A		4.0	
VIII	13	WSo	1	Mp.FC'-		(H),Sc,Cg		4.5	PHR
	14	W+	1	ma.CF.FD.TFu		Na,Fi		4.5	
IX	15	W+	1	FMa.CF-	2	(A),Fi		5.5	AG, PHR
	16	WSo	1	Mpu		Hd		5.5	PHR
X	17	DdSv/+	99	FD.ma.C-		Id,Hx		6.0	FAB2, PHR

Summary of Approach

I : W.WS	VI : W.W
II : WS.DdS	VII : DdS
III : D.D	VIII : WS.W
IV : W	IX : W.WS
V : DdS.D	X : DdS

CONSTELLATIONS TABLE

S-Constellation (Suicide Potential)

☑ Positive if 8 or more conditions are true:
NOTE: Applicable only for subjects over 14 years old.

- ☑ FV+VF+V+FD [3] > 2
- ☑ Col-Shd Blends [1] > 0
- ☑ Ego [0.29] < .31 *or* > .44
- ☐ MOR [1] > 3
- ☑ Zd [13.5] > ±3.5
- ☐ es [6] > EA [15.5]
- ☑ CF + C [3] > FC [0]
- ☑ X+% [0.12] < .70
- ☑ S [8] > 3
- ☑ P [2] < 3 or > 8
- ☐ Pure H [3] < 2
- ☐ R [17] < 17

8   Total

PTI (Perceptual-Thinking Index)

- ☑ (XA% [0.47] < 0.70) *and* (WDA% [0.62] < 0.75)
- ☑ X-% [0.53] > 0.29
- ☐ (Sum Level 2 Special Scores [1] > 2)
  and (FAB2 [1] > 0)
- ☐ ((R [17] < 17) *and* (WSum6 [11] > 12)) *or*
  ((R [17] > 16) *and* (WSum6 [11] > 17))
- ☑ (M- [6] > 1) *or* (X-% [0.53] > 0.40)

3   Total

DEPI (Depression Index)

☑ Positive if 5 or more conditions are true:

- ☑ (FV + VF + V [0] > 0) *or* (FD [3] > 2)
- ☑ (Col-Shd Blends [1] > 0) *or* (S [8] > 2)
- ☑ (3r + (2)/R [0.29] > 0.44 *and* Fr + rF [0] = 0)
  or (3r + (2)/R [0.29] < 0.33)
- ☑ (Afr [0.42] < 0.46) *or* (Blends [6] < 4)
- ☐ (SumShading [3] > FM + m [3])
  or (SumC' [2] > 2)
- ☐ (MOR [1] > 2) *or* (2xAB + Art + Ay [1] > 3)
- ☑ (COP [0] < 2)
  or ([Bt+2xCl+Ge+Ls+2xNa]/R [0.24] > 0.24)

5   Total

CDI (Coping Deficit Index)

☐ Positive if 4 or more conditions are true:

- ☐ (EA [15.5] < 6) *or* (AdjD [4] < 0)
- ☐ (COP [0] < 2) *and* (AG [3] < 2)
- ☑ (Weighted Sum C [3.5] < 2.5)
  or (Afr [0.42] < 0.46)
- ☐ (Passive [8] > Active + 1 [8])
  or (Pure H [3] < 2)
- ☐ (Sum T [1] > 1)
  or (Isolate/R [0.24] > 0.24)
  or (Food [0] > 0)

1   Total

HVI (Hypervigilance Index)

☐ Positive if condition 1 is true and at least 4 of the others are true:

- ☐ (1) FT + TF + T [1] = 0
- ☑ (2) Zf [15] > 12
- ☑ (3) Zd [13.5] > +3.5
- ☑ (4) S [8] > 3
- ☑ (5) H + (H) + Hd + (Hd) [10] > 6
- ☐ (6) (H) + (A) + (Hd) + (Ad) [3] > 3
- ☑ (7) H + A : Hd + Ad [8:6] < 4 : 1
- ☐ (8) Cg [1] > 3

OBS (Obsessive Style Index)

- ☑ (1) Dd [4] > 3
- ☑ (2) Zf [15] > 12
- ☑ (3) Zd [13.5] > +3.0
- ☐ (4) Populars [2] > 7
- ☐ (5) FQ+ [0] > 1

☐ Positive if one or more is true:

- ☐ Conditions 1 to 5 are all true
- ☐ Two or more of 1 to 4 are true *and* FQ+ [0] > 3
- ☐ 3 or more of 1 to 5 are true
  and X+% [0.12] > 0.89
- ☐ FQ+ [0] > 3 *and* X+% [0.12] > 0.89

NOTE: '' indicates a cutoff that has been adjusted for age norms.*

SCZI (Schizophrenia Index)

☑ Positive if 4 or more conditions are true:

☑ ((X+% [0.12] < 0.61) *and* (S-% [0.78] < 0.41))
   or (X+% [0.12] < 0.50)

☑ X-% [0.53] > 0.29

☑ (FQ- [9] ≥ FQu [6])
   or (FQ- [9] > FQo [2] + FQ+ [0])

☐ (Sum Level 2 Special Scores [1] > 1)
   and (FAB2 [1] > 0)

☐ (Raw Sum of 6 Special Scores [3] > 6) *or*
   (Weighted Sum of 6 Special Scores [11] > 17)

☑ (M- [6] > 1)
   or (X-% [0.53] > 0.40)

4 Total

Interpretive Hypotheses

The first step in considering the possible interpretive significance of Rorschach findings consists of determining whether a protocol is sufficiently long (more than 13 responses) and complete (no card rejections) to be useful; that is, to provide reliable data and support valid inferences. Additionally in this initial process, concerns should be raised if elevations in the S-CON suggest suicide potential. Interpretive cautions may be raised if highly unusual features of the data suggest efforts to simulate serious psychological disturbance.

This record contains a sufficient number of responses to provide reliable information and to support valid interpretations.

WARNING! This person shows many characteristics commonly observed in people who subsequently commit suicide. The possibility of suicidal tendencies and preoccupations should be evaluated carefully, and those responsible for her care should be alerted to potential suicide risk.

Constellations

PTI = 3 No	☑ DEPI = 5	☐ CDI = 1	☑ S-CON = 8	☐ HVI = No	☐ OBS =

Interpretive Search Strategy

Positive Key Variable(s)	Interpretive Search Strategy
1. D < Adj D 2. EB is Introversive	Controls > Situation-related Stress > Ideation > Processing > Mediation > Affect > Self-Perception > Interpersonal Perception

Rorschach Comprehensive System protocol

I 1. 2 birds, with their backs against each other, together, they both would like to go to different directions, they can't decide where to go, that's all I see (E: Take your time …)

E: Repeats subject's Response
S: Here are their beaks and wings, and backs, they're both going their own way
E: Where do you see it?
S: All this area

2. (v) If you turn it the other way round it's like a devil, here is the moustache, and sort of a cruel expression

E: Repeats subject's Response
S: Here are the eyes, the expression is cruel, here's the moustache, here's the head, and you can see a little of the mouth over there, you can't see all of the upper part
E: What made it a devil?
S: Just that cruel expression, it doesn't look like a human, and it's black

II 3. This is a person crying for help, the mouth is wide open, the face is kind of desperate

E: Repeats subject's Response
S: Here's a big nose, big mouth is opened, and here are the eyes

4. (v) If you turn it there's a contemplative expression, an elderly person, here's the beard, it's thinking

E: Repeats subject's Response
S: Here's its nose and here are the eyes, and here's a moustache or a beard

III 5. Here are 2 people, they have a basket between them, they both would like to go in different directions, neither gives up, and they are fighting which one gets it

E: Repeats subject's Response
S: Yes, here are the heads and the hands, here's the basket, which they both would like to have but either gives up

6. (v) If you turn it the other way it's a big fly, with its hands taking something, can't see what it's reaching

E: Repeats subject's Response
S: It's over here, here are the legs and here are the eyes, and here's the body

IV 7. Here's a motorcyclist, coming right towards, it's driving over

E: Repeats subject's Response
S: It's this, these are the legs, here's the front tire of the motorcycle, here's the head, here are the handles, it's coming right at you

V 8. Here is a meditating person, he has big moustache, a contemplating look

E: Repeats subject's Response
S: Here are his eyes, here's the moustache
E: What made it a moustache?
S: I think they just look like moustache, here's the upper part of the lips

9. (v) On the other hand 2 hands stretching, like from the sky, somewhere
from higher, can't see what they are stretching out after

E: Repeats subject's Response
S: This way here, they are stretching to something here, stretching from higher
up, whatever …

VI 10. This is at least a cross, it's on a pile of soil

E: Repeats subject's Response
S: It's over here and this is the pile of soil
E: What made it a pile of soil?
S: It's kind of a huge lump, and that cross fits into it

11. (v) Here's 2 people too, they both would like to go in different directions, they
can't leave the other, they can't give up or decide

E: Repeats subject's Response
S: Here are their heads, and here are the hands, they also have their
backs against each other, they both would like to go in different
directions, but they cannot leave anyway

VII 12. (v) An insect, I don't see anything else

E: Repeats subject's Response
S: Here are the front limbs, here the back limbs, here's the head, a big
beetle or something
E: I still don't understand how you see it there?
S: This white here in the middle is the body

VIII 13. This is a big kind of a robot-like human, with a stern look, looks like
going to do something bad soon

E: Repeats subject's Response
S: This is his head or a helmet, these are the hands, this is the body, the white
eyes here, it is like a robot because it is so wide

14. (v) On the other way round this is a volcano and soon this lava is going to cover
all that remains under it

E: Repeats subject's Response
S: This reddish is the volcano, and lava is going here on the sides, and this
lower part is land that it covers on its way
E: What made it land?
S: This is green so it is land that will remain under it
E: And what made it a volcano?
S: There's fire, and something hot that is coming

IX 15. Here are 2 dragons who fight with each other

E: Repeats subject's Response
S: Over here, it spits fire a little bit, this is their body this lower part
E: What made it look like fire?
S: It's yellow

16. (v) Then here is a man again, with a stern look, has it already done
something bad or is it going to do

E: Repeats subject's Response
S: He has beard over there, these white ones are his eyes, and these red ones
his eyebrows, he's looking under his eyebrows a little bit
E: Beard?
S: Looks like an old person, the beard fits him

X   17. (v) This is a kind of chasm that takes all the color from the world, of the so called living world, into that black hole, which narrows and narrows all the time

E: Repeats subject's Response
S: This is the world, all the color, sounds and stuff, and part is already falling into the chasm, part is on the outside
E: What made it a chasm?
S: These lines sloping downwards, a feeling that one is falling in there

138

CONTROL & STRESS TOLERANCE
  Step 1 - Adjusted D Score and CDI
  Step 2 - EA
  Step 3 - EB and Lambda
  Step 4 - es and Adj es
  Step 5 - eb

SITUATION RELATED STRESS
  Step 1 - D Score in relation to es and Adj es
  Step 2 - Difference between D and Adj D Scores
  Step 3 - m & Y
  Step 4 - T, V+3r+(2)/R in relation to History
  Step 5 - D Score (if appropriate Pure C, M⁻, M no form)
  Step 6 - Blends
  Step 7 - Color Shading Blends

AFFECTIVE FEATURES
  Step 1 - DEPI & CDI
  Step 2 - EB & Lambda
  Step 3 - EBPer
  Step 4 - Right Side eb & variables related to it
  Step 5 - SumC':WSumC
  Step 6 - Affective Ratio
  Step 7 - Intellectualization Index
  Step 8 - Color Projection
  Step 9 - FC:CF+C
  Step 10- Pure C
  Step 11- Space responses
  Step 12- Blends (Lambda & EB)
  Step 13- m & Y blends
  Step 14- Blend complexity
  Step 15- Color shading blends
  Step 16- Shading blends

INFORMATION PROCESSING
  Prerequisites (L,EB,OBS,HVI)
  Step 1 - Zf
  Step 2 - W:D:Dd
  Step 3 - Location Sequencing
  Step 4 - W:M
  Step 5 - Zd
  Step 6 - PSV
  Step 7 - DQ
  Step 8 - DQ Sequencing

MEDIATION
  Prerequisites (R,OBS,L)
  Step 1 - XA% & WDA%
  Step 2 - FQnone
  Step 3 - X-%, FQ- frequency, S- frequency
  Step 3a- Homogeneity issues
  Step 3b- Minus distortion levels
  Step 4 - Populars
  Step 5 - FQ+ frequency
  Step 6 - X+% & Xu%

IDEATION
  Step 1 - EB & Lambda
  Step 2 - EBPer
  Step 3 - a:p
  Step 4 - HVI, OBS, MOR
  Step 5 - Left side eb
  Step 6 - Ma:Mp
  Step 7 - Intellectualization Index
  Step 8 - Sum6 & WSum6
  Step 9 - Quality 6 Spec Scores
  Step 10- M Form Quality
  Step 11- Quality of M responses

SELF PERCEPTION
  Step 1 - OBS & HVI
  Step 2 - Reflections
  Step 3 - Egocentricity Index
  Step 4 - FD and Vista (in relation to History)
  Step 5 - An+Xy
  Step 6 - Sum MOR
  Step 7 - H:(H)+Hd+(Hd) & Review codings
            for Human Content responses
  Step 8 - Search for projections in:
  Step 8a- Minus responses
  Step 8b- MOR responses
  Step 8c- M & human content responses
  Step 8d- FM & m responses
  Step 8e- Embellishments in other responses

INTERPERSONAL PERCEPTION
  Step 1 - CDI
  Step 2 - HVI
  Step 3 - a:p Ratio
  Step 4 - Food responses
  Step 5 - Sum T
  Step 6 - Sum Human Contents & Sum Pure H
  Step 7 - GHR:PHR
  Step 8 - COP & AG frequencies & codings
  Step 9 - PER
  Step 10- Isolation Index
  Step 11- Contents of M & FM responses with pairs

APPENDIX 3 Interview Formula

What kind of work are you doing right now?

How long have you been doing clinical work?
- In a hospital ward?
- In an outpatient clinic?
- As a psychotherapist?
- Something else?

What other kind of work have you been doing as a psychologist?
When did you graduate from the Rorschach Comprehensive System education?
How long have you used Comprehensive System in personality assessment?
About how many times?

How much / how often have you done assessments?
- In hospital wards?
- In outpatient clinics?
- As a psychotherapist?
- Something else?

For what kind of purposes have the assessments usually been done?
How long have you used WAIS in assessment?

Basic education (PhD)

Graduated what year:
Postgraduate education 1:
Graduated:
Postgraduate education 2:
Graduated:
Postgraduate education 3:
Graduated:
Postgraduate education 4:
Graduated:
Postgraduate education 5:
Graduated:
What kind of theories or ways of thinking are most familiar to you?
Email address:

APPENDIX 4 Certification on Confidentiality

The undersigned researcher, a clinical psychologist working in the Tapiola out-patient clinic of Jorvi hospital, is doing a doctoral thesis in the University of Jyväskylä, about how clinicians think when they do personality assessment. The research is supervised by professors Carl-Erik Mattlar, Jarl Wahlström and Pertti Saariluoma.

The research studies the ways clinicians think when they do personality assessment, and the identities of subjects are not relevant. Their identities will not be revealed in any phase of the research, and the research will be carried out in a way that the anonymity of the subjects will be completely protected. Only the undersigned will get to know the subject's identity while gathering the research material, but no one else. Not even the supervisors of the research know who they are. The anonymity will be treated according to normal professional ethics in psychology.

Practically this means that after the researcher has taped the assessment process, all the material will be numbered, and referring to the material is done only by using these numbers. The taped protocols will not be referred by initials, by short characterising or in any other way but using these numbers. In the publications there can be some fragments of protocols, but the subjects will not be identified from these fragments.

With this signature I respect and fortify the above described confidentiality.

Olli-Pekka Santala
Psychologist
Tapiola outpatient clinic/
Jorvi hospital
Opinkuja 2
02100 Espoo
Finland
Tel work +358-9-861 7650
Tel home +358-40-734 6561
Fax +358-9-861 7670
email: op.santala@saunalahti.fi

APPENDIX 5 Protocol 37A

The Free Assessment Phase

P = Participant.
E = Examiner.
Bold text = A verbalization of the participant that includes something else than what existed in the materials.
? = Unclear word.
… = Short pause in talking.

P: *Okay 24-year old male, er high school er comes for depression, young man whose depressed, er no previous treatment, some years ago when his parents er parents divorced* **okay**, **kind of** *financial trouble, and gave up his savings to help them,* **okay** *and he had been working and saving money to buy himself an apartment, and bought a BMW, when his mother told him about the parents financial difficulties and that their rents would be distrained, he took all the money from his savings account, paid some of his parents debts and worked in two shifts to help them. At the same time his friend totally crashed his car,* **oh my goodness**, *so he lost both his savings and his car. His financial situation was bad and he gave up his apartment and moved in with his mother,* **okay**, *… he says that he had a nervous breakdown, to relief his bad feelings he started to drink,* **self medicating the depression**, *drinking increased so much he could not always take care of things, drunk driving,* **okay**, *he could stop drinking by going to a short treatment in a clinic for alcoholics and getting appropriate psychiatric medication, he has now been on a medical leave for more than a year* **uuuh**, **doesn't say what kind of work he does**

E: *It'll come later on*

P: **Okay**, *mostly spends time home at his mother's place, without seeing people, doing nothing special or sometimes watching videos, mother does all the homework, ashamed of his situation that he cannot go out, no telephone and doesn't open his mail, mother gives him the most important letters, tells he gets anxious very easily, gets bad news through mail, an invitation to a trial or a larger bill, he gets anxiety attacks, on the bus he may get panicked, chronic insomnia,* **okay**, *uses sleeping pills and tranquillizers more than doctor ordered, is the younger of two children, his sister that is older, works in an office has a boyfriend, his parents divorced 4 years prior,* **so he was 20**, *mother is from Russia and unemployed, father is an alcoholic and has been violent,* **bad relationship**, *father has severe mental problems, but not been in any treatment, and has long been without apartment since the di… divorce, occasionally the father visits the patient and ex-wife, probably to get s… stay gets to stay inside overnight, and he often goes into one of the rooms and sits in the dark without saying a word for a couple of days until he leaves, patient thinks these are distressing situations,* **I agree**, *doesn't want to talk to his father, and when his father rarely says something criticizes his son, patient went normally through comprehensive school and high school, achievements were average, in the beginning of high school he felt things were going pretty well, he had some hobbies, taekwondo, badminton, fishing and weight-lifting, in the end of high school he had to work to get some money, as a guard and a taxi driver, financial disagreements in the family, they couldn't afford to buy him school books, but then also he himself did not want to use his earnings in buying books, high school time was difficult, and sometimes he worked around the clock, after high school he had some short employment's in a warehouse, as a guard and driving a taxi, he did his compulsory military service in 2001, sa… he was then in very bad nutritional condition, gained 20 kilos, life was like a vacation in the army, he had some distance from the father, after the army worked 3 months in a warehouse,* **how long was he in the army? One year?**

E: *Yeah about er eleven months*

P: **Eleven months**

E: *Yeah*

P: **Okay**, *he thought he would try to get to a technical school, but he couldn't go to the entrance exam because of his symptoms, when he is at the psychologists office his hands are wet, he is so tensed, but in spite of this he tells things in an organized way, openly and good cooperation, no suicidal thoughts, no impression of any psychotic features, constantly in a low mood and helpless about his future, feels everyday routine things are overwhelmingly difficult and normally just stays at home,* **all right lets see what we got for intellect, average intellect, mainly consistent with academics although good** *similarities, picture arrangement* **looks good**, *information is* **low**, *digit span* **he's got some problems with concentration**, *vocabulary is* **low, so maybe there might may be a little better, let's see what his R looks like**, *19 Responses, PTI is* **okay**, *DEPI is* **okay**, *CDI is* **okay, the rest** *mmmm, SCON,* **not** *Hypervigilant,* **not** *Obsessive,* **well let's see how he approaches this**, *sequence of scores,* **keeping it simple**, *… okay, picks up his minuses on card 10, also gets his Fab on card 10, well we've got a* **high** *L, … hmmm D and AdjD are positive +1,* **maybe misleading, yes**, *FM is* **low**, *thinking, what is his Special scores, DV FAB one FAB2, Xu% is* **pretty high**, *hm,* **many** *D vagues,* **I don't like that**, *one Pure C, Afr is a* **little low**, *3 White space* **not too bad**, *GHR PHR is* **not great, surprise that he's not**, *active to passive* **isn't greater**, *… hm* **low** *Egocentricity index, Pure H to non Pure H* **while low it's in a good range, a good** *?, Zd is a +2* **normal range, okay** *SCON … DEPI got the 3 White space,* **low** *Egocentricity,* **backs off from affect, okay I'm ? away of** *COP, Hypervigilance,* **don't expect much from people**, *with his* **history that would be consistent, nothing** *on the PTI,* **surprise the CDI isn't higher, close on that one, close on that one, … close on that one, so he's actually close on several of them**, *… SCZI* **that's a nonissue**, *… all right, Interpretative search strategy is what I want, what I expect, a bug a butterfly bit of a feeler wings here's the middle body, could be a leaf that is a little bit ripped, that's a Morbid, edges are ripped holes in the middle ?, er nothing looks a little bit like a butterfly hmmm,* **okay**, *these could be the wings a little bit like a head this black area, so he's on card 2 there, D6, well, something that is ripped in the middle hmmm, ripped ripped, ripped I don't know there's some hole or then that butterfly just has a white splash over there, inquiry, here in the middle, ripped, I don't know some hole,* **how is that coded for**, *… two persons sitting together what are they doing something in the middle and a little bit like in that show ten thousand dollar questions computers or something kind of facing each other ?, some bug legs teeth eyes uuup bugs with teeth where you see it, the upper part,* **all right now we have bug again**, *these could be like legs I'm not sure what makes it look like that,* **and here we got leaf again, we got the** *leaf* **back here** *in 2,* **another** *butterfly, Response ten, leglers,* **how that was coded**, *plant, two continents, could be some kind of mainland somewhere,* **he is in another world all right**, *a mushroom and ink has been thrown on it* **well that's a good answer**, *there's a flower variegated,* **several** *flowers this one is a rose, coat of arms, over here could be a person flying here are the wings it's quite a sparkling picture a psychedelic figure, looks like two electrical discharges, yes, and fireworks ? up, some part of a rocket banging colorful,* **okay well I got a young er what would you like me to do, kind of put together my thoughts about it now or**

E: *You can first think about it if you want to but at some phase I ask you to sum up your view*

P: **Okay**

E: *It depends how you work, you can do it right now or you can just think about it for a while before we go into that*

P: **I thought I could just think about it out loud I got a young man who was fairly average but clearly family turmoil** *er* **has been part of his life for a long time**, *er* **it sounds like** *er* **when he is more structured and supportive he, supported he does quite well**, *er* **he's in high school and he's got good support he does okay, when he is in the military got support** *er* **he does okay and** *…* **he does better when he's got some structure** *er by* **history**, *er* **he has had a** *er* **kind of a poor foundation** *er* **in in the past, in in in growing up with this turmoil** *er* **in the family**, *er* **I do have some concern when reading the Responses, that there is** *er* **a bit of perseveration that goes through as I'm reading at this is not a a specific coded** *er* **part of the R but at that the** *butterfly* **which is** *er* **perseverated and that** *ripped* **which is perseverated I got** *bugs* **which are perseverated I got** *leafs* *er* **which are perseverated** *flowers* **which are perseverated** *er* **and that makes**

me again think of some real difficulty generating options in his life, he's got some er *he kind of sticks with something even though it's not working for him,* er *he's he gets pretty occupied by something,* er *and so I wanna think about that some as I go through* er *he's got 2 M's and he's got 1 M-,* **I want to look at that** M- **again, I wanna see what that is, it's gotta be this one** *the man flying 17, flying person, here are the wings a bit of leg here's the head it's a sparkling picture, lot of colors has been thrown in there, kind of psychedelic figure, er,* **... in general when I when I think about him his** er **he has some ability if I look at his** er **intellectual testing to think abstractly with his** *similarities being up,* **he has some idea about** er **sequencing and and appropriate social relationships because of his** *picture arrangement* **but again those seem to fail him to some degree in the interpersonal relationship, he is somebody that is withdrawn,** er **from the relationships, doesn't understand people very well, so I think it's been easy to withdraw,** er **the Self perception is** **pretty low,** er **and and I'm quite concerned about** *the DQ vagues,* **and about what I see as** er **this repetitive nature of the content,** er **it makes me wonder** er **one has he had a head injury,** er **because I would be I would be concerned about that, if he'd had a head injury** er **then** er **I would have a better understanding of it, if he hasn't had a head injury then I have somebody who is** er **fairly primitive in the organisation,** er **and and is likely to to do the kind of thing that he's doing, either role of the son, the good son with the mother,** er **to withdraw to use that, as a way not to have to go out into the world and deal with things,** *I think he's D+1 and AdjD+1 is is* **not really accurate,** er **he's not placing himself in a position of experiencing er much from the world,** er *the FM being 0 er it* **doesn't mean that there aren't any need states, it basically means that he's not allowing them to build up, he withdrawing isolating himself, keeping himself** er **away from interacting with other people, keeping it quite simple, does not wanna look at any complexity in the world, it's too much for him, overwhelms him,** er **he doesn't have a good way for dealing with that, I'm very concerned about** er **both of the familial potential for substance abuse based on** *his father,* er **I'm concerned** er **about a** er **potential for inappropriate acting out based on** er *the DQ+* **and if he's overwhelmed by affect and if it is coming out now,** *and* **anxiety,** er **withdrawal i... is protecting him, but I think if he's forced to be out he may not do as well, I think** er **what I would like to see for him moving to treatment is maybe a a longer term** er **very structured program that moves him from** er er **and I'm not talking about a hospital I'm talking more about a program that focuses on developing skills, maybe some group therapy, but also focuses on vocational achievement, gives him some very clear guidelines to begin to develop a sense of independence,** er **I would also think that** er **that program needs to be** er **structured so that he will make progress and feel positive about himself, I think individual therapy aimed at helping him understand himself to some degree and helping him understand feelings to another degree** er **would be very appropriate, I think his problems with** *insomnia er we see a little m of 2 it* **doesn't surprise me,** er **I often will find people who're having difficulties with sleep with that** *little m or 2,* **I think that will resolve I'm not too worried about that,** er *antidepressant medication er* **I don't see clear evidence for a need, I don't know what kind of medication he's on,** *tranquillizers* **I'm concerned about that, concerned about** *tranquillizers,* er **I think if he has been in** *taekwondo* **in the past I'd like to see him do that again, I think that kind of active movement and structure is important, let's see he doesn't say anything about a spiritual base** er **I might explore that and again a positive spiritual experience might be helpful for this young man as an overall** er **intervention, ... he very much more is likely positive CDI than not, very much more is like that and he's close to that and and that the interventions with him will be more like a patient with a positive CDI**

E: *Okay*

P: **All right that's where I would think through it**

144

The Description Phase

P = Participant.
E = Examiner.
? = Unclear word.

E: *Could you now sum up your view on his personality, don't speak so fast I have to er write it down*

P: *Sum up his personality*

E: *Yeah*

P: *Er this is a young man whose history er has er predisposed him to difficulties, who has done better when in structured era setting, when the structured setting has been taken away from him through family divorce he tries to fill the role of providing for mother who is the victim er but he is inadequate to do that*

E: *Excuse me?*

P: *Is inadequate to do that, based on his inability to stay with his education and probably making some bad judgments about his er lifestyle, er so th… so there is some judgment problems, his thinking is er somewhat er immature and er rather unsophisticated, he has chosen to deal with the problems in his life by withdrawal and isolation, … there's is no clear depression, there's no clear thinking disturbance, er but there is this primitive quality to his makeup that is likely to er undermine him in the future*

E: *You said primitive quality in his?*

P: *Psychological makeup*

E: *Ahaa*

P: *… I'm not sure how much more to say, yeah I mean I think I have*

E: *It's up to you*

P: *I think I've said about what I would like to see in terms of treatment, what I would like to see in terms of treatment is structure er in his life, er I always ask myself do I like this person, do I not like this person, er I always ask myself do I want to work with this person*

E: *er you don't to have to er think about the treatment now if you don't want to, it's enough to think about the personality*

P: *well that's what I'm trying to do when I think about er you know do I do I is this somebody that I'd think I would like to work with because I like his personality, I like something sort of there, or is this something that I think would be difficult to work with, er should not be an easy person to work with this is going to be longer term er because of of having some primitive personality makeup, I also am concerned er about the family history and about the drinking that he is doing and that will make things more difficult, er and I guess from my perspective er he would be okay to work with, he would be er I don't have to put a lot of energy into working with him, so er I'm not sure if I wanna work with him or not, I'm not sure how much I like him, he's okay I guess but I don't know that I'd wanna work with him right now*

The Inquiry Phase

P = Participant.
E = Examiner.
Bold text = An item from the materials that has been used.
? = Unclear word.

E: *Okay, I got to ask you a couple of questions*

P: *Sure*

E: *You mentioned er judgment problems, on what do you base that idea*

P: *I'm concerned that I see he has **2 M's** and he's got **1 M-,** there is some judgments issues that go along with that, … and and I think there's there's a fantasy quality when you think of the **content** er of the **one good M**, and that is they're doing the seven or the **ten thousand dollar questions**, but nothing's happening, that nothing is going on, so there's a fantasy quality, so I think his judgment at times will be clouded, his understanding of people er is not er particularly good er so I think judgment around people will be difficult*

E: *On what do you base that idea?*

P: ***GHR PHR** er the **need to er rush in and take care of mother** er while at the same time **allowing father to come back and spend time**, er there's a these are my k… my thinking about the judgment, and the u… and the using of the **alcohol**, bad judgment*

E: *Yeah, how about er immature and unsophisticated thinking, may be a little the same thing but sounds a little bit different too?*

P: *Well the immaturity and the lack of sophistication comes in both with the **high L** and the **high number of vague Responses**, er and in the more qualitative way is the repetition of the **content** that I that I see throughout the protocol*

E: *How about the idea of isolation and withdrawal?*

P: *Er I like that the testing is consistent with what the history shows, er he clearly has er **retreated into this apartment with his mother**, er the **Isolation index** is positive, er so it it's consistent, so I like that*

E: *You mentioned that there is no clear depression, how come, where do you base that?*

P: *The the **Depression index** is is low, er my experience has been when there's somebody at the level of **4** that are not going to er do as well with **antidepressant medication**, er so so I don't see it clearly in the **R**, er more what I see is is a **withdrawal from engagement** er and more a fear of having to produce of having to engage er the the **insomnia** er that he talks about well that might be an indicator of depression and I think it's more an indicator of rumination about his situation and not knowing what to do with it, just feeling trapped, so not not not a clear depression but more that er lack of ability to cope adequately*

E: *How about no clear thinking disorder*

P: *When we do get some problems with his thinking which was the er I think that a **Fabcom2**, er as that comes in er on **card 10** er and that really is more related to that affect and that lack of er good grounding with people er er and I think the affect overwhelms him er and and causes his thinking er to to be modifying, so I again I don't see it as a er schizophrenic kind of er thinking problem*

E: *Yeah, so you mean there is a thinking problem sometimes which is related to affect but it's not very much?*

P: *Well it can be very much when the affect is evoked, if there's more affect evoked then the thinking I think will get worse cause he wants to withdraw that's what withdrawal is about, er so I think er more affect there is the more problems with thinking there will be now*

E: *Okay, then you mentioned that there is a primitive quality in his psychological makeup, what do you mean by that?*

P: *Well I think I've identified that a little bit with the **DQ vagues** and the **high L**, er it's kind of oversimplifying, it's it's a an immaturity er of of how he deals with things er that's that's mostly it*

The Phase of Explaining or Understanding

P = Participant.
E = Examiner.
? = Unclear word.

E: *Okay, one last question on this case, and so again a little bit overlapping, but how would you no… now in light of this personality picture you made er explain or understand his situation, shortly?*

P: *… I think that the way that I would explain the situation is er the the issues for maladjustment or the foundations for maladjustment have been there for a long time, the situation of father and mother divorcing allowed that maladjustment to escalate, the mother allowing the child to to take care of her, allowed that maladjustment to escalate er he I think he was going to have some difficulties regardless of that happened or not, the lack structure for being out of school allowed that maladjustment to escalate so I think there w… ? (268) I think we are looking at a developmental issue, developmentally he was not strong when he became a young adult and I think something was going to happen that was not going to be positive and was would require treatment*

E: *Okay, thank you.*

APPENDIX 6 Instructions for Coding the Content of Personality Characteristics

General coding principle: The whole verbalization is coded, not just one word, although coding is based on words.

Cognitive characteristics (COG)

Definition: Descriptions of thinking, reality testing, ways of perceiving, information processing, etc.

Examples: *Is not realistic, a tendency to misperceive situations, disorder that disturbs thought process, difficulty differentiating what's important from what's not, doesn't understand people well, perception of the world is quite distorted, does not see the world very accurately, even if she could see the world accurately she's really unique in how she perceives the world, suicidal thoughts, problems with reality testing, anger causes the distortion of reality testing, distorted perception, strong feelings of anger disturb reality testing, difficulties in perceiving relationships, doesn't like to process strong feelings.*

Neuropsychological performance (NP)

Definition: Descriptions of cognitive performance, like intelligence, visual or language performance, etc. Usually all kinds of descriptions based on neuropsychological tests and expressed in neuropsychological terms. Descriptions are coded COG prior to NP, for example, *is very intelligent in verbal thinking* would be coded COG because it's about thinking. Also, *has capacity for more complex thinking and processing* would be COG.

Examples: *Talented, is cognitively talented, can't really abstract the information that she has access to, cognitive capacity okay.*

Experiental characteristics (EXP)

Definition: Descriptions of what the person experiences, such as *depressed, angry, needy, empty, sensitive, confused, threatened, hopeless, overwhelmed, alienated,* etc. Also *emotion, trust, feeling lost, overstimulated, wish, hostile, identifies with real people* would be EXP. This class describes experiences no matter what the content of experience is, such as in *suffers from feelings of emptiness* or *confused about other people.* However, *emotions are not well integrated* would be GEN, since it does not describe any experience but more how the personality is structured. Also, descriptions of needs is coded GEN. Still, *tendency to be depressed* would be GEN, as it talks about tendencies.

Examples: *Experiences that the environment is bad, very angry, suffers of feelings of emptiness, suffers of depression, is in an internal chaos about things, very needy, de-*

structivity and aggression are problematic, her trust has been traumatized, acute sensitivity to abandonment, enormous amount of confusion about other people, experiences other people as threatening, feels extremely vulnerable, hopeless, depressed, her way of interpret things confuses her, at times thinking about the situation overwhelms her, feels over stimulated, needs safety and support, when affect is evoked it is angry and distorted, hopes that someone would make her life easier, anxious, feels bad about things, hostile, very angry, affects are loose and labile, depression is about emptiness, alienated, desperate, big problems with aggression, feels that problems are outside herself, low self-esteem.

Relationship to self (SELF)

Definition: Descriptions of a person's relationship to himself/herself. SELF is preferred to COG, EXP and GEN.

Examples: *The way she sees herself is based more on fantasy than reality, there's some self-reflection, does not value herself.*

Motivational characteristics (MOT)

Definition: Explicit descriptions of motivations to internal or external action.

Examples: *Strives to be independent, tries to protect herself, not motivated to psychotherapy.*

Coping characteristics (COPE)

Definition: Descriptions of actions with which the person is seen as reacting to or acting against different kinds of demands in life. The demands and respective actions may be internal, as in defensive or problem-solving activities, or external, as in behaviors that target at solving a problem. Also descriptions of resources, except descriptions of intelligence or cognitive features, which would be coded NP and COG. "Coping" may also be translated as actions targeted at controlling, regulating, etc. It must be noted, that this way to conceptualize "coping" deviates from its traditional meaning.

Examples: *Projects the bad ideas of self into the outside, resources go into trying to control bad feelings of self, is not able to use his talents to manage in life, more considers things in her mind, problems with anger, difficulties in modulating affects, the feelings of loneliness and being abandoned make her strive to actualize some kind of relationship, uses an enormous amount of energy to ruminated herself but it is not successful, not sure she can bring about relationships in a way that she wants to, personality is organised to try to understand and fend off the bad feelings of being damaged, emotional control is not good, distorts to a level that will disrupt her functioning, there are resources but is not able to use them, difficulties in maintaining a continuing and stable feeling of*

self, has not been able to grow up, uses processing as a shield, a lot of resources, a tendency to escape to fantasy.

General personality characteristics (GEN)

Definition: Any general description of personality, which at the same time would not be coded COG, NP, EXP, SELF or MOT. Tendencies and dispositions are coded GEN.

Examples: *Not traumatized, not separated, very childish, her mind is very alive, a long-standing disturbance of personality, the risk of suicidality comes from the dynamics in personality, sexually confused, disintegrated, weak ego, no stable identity, very sensitive to react, very rigid inner logic, inner logic is based more to fantasy than reality, dissociated, problems in identity, unstable, fragile, immature, flexible, the development of personality is not finished, prone to be psychotic, vulnerable to emotional stimulation, stressed.*

Behavioral characteristics (BEH)

Definition: Descriptions of external behavior. Descriptions are coded BEH unless they are coded COPE. Also, dispositions as in *may react as if she has been cruelly cheated or betrayed* is GEN, not BEH.

Examples: *Can act in a very aggressive way, possible self destructive act would be to send hostile message and embarrass someone, impulsive.*

Relatedness (REL)

Definition: Descriptions of relations with other people. If description talks about how the patient experiences other people, this is coded EXP, not REL.

Examples: *Seeks for safety of relationships, doesn't have good relationships, tries to compel the outside world into her own inner logic, stucked into symbiotic way of being with people, passive in relationships, isolated, does not act conservatively in social situations.*

External circumstances (EC)

Definition: Descriptions are coded EC when they describe background, environment, situational factors, etc.

Examples: *Traumatic background, life situation has brought her to the brink of disaster, has been hurt a long ago.*

Diagnostic characteristics (DG)

Definition: Explicit descriptions of psychiatric diagnoses.

Examples: *Not schizophrenia, more depression than schizophrenia, mood disorder.*

Phase diagnosis (PDG)

Definition: Description of phase of diagnosis or level of diagnosis, like neurotic, borderline or psychotic. Also if some level of personality functioning is described using phase diagnostic terms. If a description is not PDG , it will usually be coded GEN.

Examples: *Borderline level, borderline organization, psychotic episodes are possible, mostly functions at the low level of borderline, moving into psychotic end of it, depression can be psychotic.*

Treatment characteristics (TR)

Definition: Description of treatment.

Examples: *A long treatment and maybe in a group, her very distorted thinking could be improved if she could begin to get the experience of safe relationship or set of safe relationships with some sense of belonging, some of the newer anti-psychotic medications might help.*

Other characteristics

Definition: Descriptions that have been difficult to classify in any more specific way.

Examples: *Very real risk of suicidality, suicidal, can be suicidal.*

SOME DECISION MAKING ISSUES

EXP versus REL: *Enormous amount of confusion about other people, experiences other people as threatening, interested in people, misses people, angry towards others, hopes that someone would make her life easier where she has not succeeded herself, experiences social situations as difficult, closeness is difficult, paranoid, needs people* would be coded EXP, as they are about how other people are experienced, but *seeks for safety in relationships, doesn't have good relationships* would be coded REL. Also *tries to compel the outside world into her own inner logic, isolated,* would be REL, but *alienated* would be EXP.

EXP versus TR: *Needy* is EXP, but *needs support in everyday life, needs social and rehabilitating support, needs help in evaluating things* are TR.

EXP versus GEN: EXP is preferred to GEN.

EXP versus COPE: *Angry* would be EXP, but *problems with anger, difficulties in controlling emotions* would be COPE, so COPE is preferred to EXP.

EXP versus COG: *Is unrealistic in interpreting things* would be COG, but *the way that she interprets things confuses her* EXP. *A lot of fantasies of her life* would be EXP, not COG.

EXP versus SELF: *Low self-esteem* would be EXP, but *does not value herself* SELF.

EXP versus DG: *Depression* would be EXP, but *more depression than schizophrenia* would be DG.

COPE versus COG: COPE is preferred to COG. Therefore, *a style where things are solved by thinking, there is something strange and illogical and immature and impressionistic and unsophisticated in the problem-solving style* would be COPE. But, *as a result of anger there's distortion of reality testing, thinking gets distorted, makes bad judgments of his life*, would be COG.

Is rigid in thinking and behaving might theoretically be either COG, COPE or BEH, but it would be coded COPE.

COPE versus GEN: *Weak ego* would be GEN. *Extratensive, extratensive character*, would be GEN, but *extratensive problem-solver* COPE. *Narrow* would be GEN, but *narrows down* COPE.

COPE versus REL: *Difficulties in building and maintaining long lasting relationships, weak social skills, difficulties in social relationships*, would be COPE, so COPE is preferred to REL.

COPE versus BEH: *Ability to work and function* would be COPE, so COP is preferred to BEH.

COG versus OTHER: *Suicidal, suicide risk* would be OTHER, but *suicidal thoughts* COG. Similarly, *wants to harm himself* would be EXP, not OTHER.

GEN versus REL: *Dependent on other people* would be REL, but *dependent* GEN.

MOTIV versus TR: MOTIV is preferred to TR, so *not motivated to psychotherapy* would be MOTIV.

DG versus PDG: *Borderline personality* would be PDG, not GEN, and *borderline personality disorder* would be DG. *Psychotic depression* would be PDG, not DG. So always when a phase diagnostic term is used, PDG is preferred. Exceptions are descriptions of tendencies, like *a tendency to be psychotic based on the trauma*, which would be GEN.

BEH versus EXP: *Loose in expressing emotions* would be BEH, not EXP.

NOT POSSIBLE TO DECIDE

Sometimes the proper classification is impossible to decide. Mostly, this seems to be for one of three reasons. First, some characteristics are too unspecific to give any basis for classification, such as in *at the present time very unstable*. Second, some are too specific and cover too many angles to give any rationale for one specific decision, such as in *trauma is playing itself out in relationships with these married women and these fantasies and desires towards them*, which might be both REL and EXP. Third, some expressions naturally denote more than one possible interpretation. For example, in *how depressed, how bleak, how hopeless the world looks to her* the word *looks* may be taken to mean either how the patient experiences the world or how the world is perceptually interpreted.