THE PROCESS AND OUTCOMES OF DIAGNOSTIC PROBLEM SOLVING AMONG EIGHT READING CLINICIANS

By

Ethelyn Maxwell Hoffmeyer

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Elementary and Special Education

1980

ABSTRACT

THE PROCESS AND OUTCOMES OF DIAGNOSTIC PROBLEM
SOLVING AMONG EIGHT READING CLINICIANS

By

Ethelyn Maxwell Hoffmeyer

If correct diagnosis of reading problems is a prerequisite of effective remediation, then one approach toward improving remedial practices might be to improve the diagnostic skills and training of reading clinicians.

One effort to study clinical diagnosis in medicine and subsequently in reading has been research relating to the Inquiry Theory of Clinical Problem Solving. The Inquiry Theory was developed by a team of researchers at Michigan State University to provide a formal theoretic structure that would integrate and account for the numerous concepts and empirical findings on clinical problem solving.

The major objective of this study was to answer the following questions relative to the Inquiry Theory:

1. Do experienced reading clinicians agree on the data they collect for a specific reading case in order to make a diagnosis?

2. Do experienced reading clinicians agree on the diagnostic statements they make for a specific reading case in order to make a diagnosis?

3. Do experienced reading clinicians use hypotheses to direct their inquiry in diagnosis?

Eight experienced reading clinicians from the mid-Michigan area participated in this study. Each clinician in three clinic sessions, no less than one nor more than four weeks apart, interacted with three simulated reading problem cases. The third case, unbeknown to the clinicians, was a replicate of the first case. The four simulated cases and their equivalent forms used in the study were counterbalanced to minimize systematic effects. Clinicians were randomly assigned to case order.

Procedures were as follows. Clinicians were asked to:

1. interact with materials of a simulated reading problem case,

2. write a diagnosis and remediation,

3. transfer the written diagnosis to the Reading Diagnostic Check List,

4. check responses to questions concerning _why_ they asked for certain case data and _what_ information the data provided,

5. indicate the content of a "good" diagnosis,

6. explain how they usually conduct a diagnosis, and

7. define for a specific case the skills of (a) instant word recognition, (b) word analysis, (c) reading fluency, and (d) reading comprehension.

To measure subject reliability on use of the Reading Diagnostic Check List, the clinicians were mailed an uncircled, carbon copy of each of their written diagnoses one week after the third clinical session. An accompanying letter instructed subjects to follow the same procedures they had used in the clinical session for transferring their written diagnostic statements to the check list.

Analysis of the data consisted of (1) formal product measures (including proportional agreement, commonality scores, inter/intra- and intraclinician agreement Phi correlation, and the Porter statistic), (2) formal process measures (using correlation, partitioned Phi coefficients, and cue-to-statement relationship statistics), and (3) informal product/process measures (using Sherman's Model of Reading and Learning to Read).

Regardless of the small sample of reading clinicians participating in this study, there is evidence to support a number of conclusions. These conclusions are:

1. Experienced reading clinicians using simulated reading cases appear not to share a common data base (memory) regarding what information (cues) should be included in a diagnosis or what diagnostic statements are important in writing a diagnosis.

2. Experienced reading clinicians using simulated reading cases appear not to share a common diagnostic routine (strategy) in terms of how to go about a diagnosis.

3. Experienced reading clinicians using simulated reading cases appear not to use consistently a theoretic process model of reading diagnosis as might be reflected in hypothesis-directed inquiry.

This dissertation is dedicated with admiration and appreciation to Dr. Byron VanRoekel. It was he who first sparked my interest in the reading profession. His love and concern for children, his understanding of the many facets of the reading process, and his interest in sharing his skills to teach reading teachers and clinicians have been an inspiration to me throughout my graduate school and teaching years.

To Dr. Van Roekel, a heartfelt thank-you for the many years of dedicated work with the children and reading specialists who were and are a part of the Michigan State University Reading Clinic.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to those individuals who have been an inspiration and a help to me throughout this research endeavor:

Dr. John Vinsonhaler, who directed my research, for his confidence in my ability and for sharing his many research skills and his enthusiasm in the Clinical Studies Project.

Dr. George Sherman, my doctoral committee chairman, for his faithful support. His ability to communicate his knowledge and understanding of the reading process helped me develop many ideas that went into this dissertation.

Dr. Lawrence Lezotte, doctoral committee member, for his interest in my work and his insightful suggestions. His sunny disposition and encouragement helped me forge ahead and complete this study.

Dr. Glen Cooper, doctoral committee member, for his interest and his helpful comments.

Dr. Donald Freeman, for his scholarly and creative editorial comments. His suggestions were invaluable in the completion of this dissertation.

Dr. Andrew Porter, who developed the research design for this study and who gave unselfishly his valuable time and expertise.

Mrs. Ruth Polin, of the Clinical Studies staff, for her patient assistance with the computer terminal and the data analyses.

Mr. Christian Wagner and Ms. Francine Kitchen, of the Clinical Studies staff, who designed the computer programs used for data analysis and assisted me in those analyses.

Dr. Lois Bader, for sharing her expertise in reading and helping me formulate many ideas for research.

Dr. C. Jay Stratoudakis, for her help in piloting the procedures used in this study.

Mrs. Jerri Zemper and Ms. Debie Salters, for their technical assistance in preparing the dissertation proposal.

Dr. Keith Goldhammer, Dean of the College of Education and my favorite teacher, for providing me with the incentive to strive for excellence.

Mrs. Susan Cooley, for her professional expertise in editing and typing this manuscript. Her organization and her ability to work calmly under pressure are amazing.

My husband, Tom, our children, Kevin and Kelly, and my parents, Mr. and Mrs. Ralph Maxwell, for their sacrifices, their tolerance, and their encouragement.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	xiii
LIST OF APPENDICES	xiv
Chapter	
I. THE PROBLEM	1
Introduction	1
Purpose	4
Theory	5
Introduction	5
The Clinical Encounter	6
The Basic Management Information System (BMIS)	14
Theoretic Implications	15
Summary	20
Overview	21
II. REVIEW OF THE LITERATURE	23
Introduction	23
Research Procedures in Problem Solving	24
Total-Task Studies	24
In-Basket Studies	26
Tab-Item Methods	27
Problems Inherent in Reading Diagnosis	28
The Essence of Reading Diagnosis	28
The Debate Over Causation	35
Lack of Standardized Terms	37
The Inquiry Theory: Related Research	40
Introduction	40
The Medical Inquiry Project	41
Clinical Information Processes in Reading (CLIPIR)	43
Observational Study, 1977	45
Outcomes, 1977	45
CLIPIR Application Exploratory Study in Educational Research (CAESER), 1977	46

Chapter		Page

Implications, 1976-1977 46
Recent Studies 47
Summary 47

III. DESIGN AND PROCEDURES 50

Introduction 50
Sample 50
Research Design 51
Instrumentation and Data Collection 54
Reliability 55
Instruments 59
Statement Concerning Hypotheses 60
Procedures 61
Summary 64

IV. ANALYSIS MEASURES AND RESULTS 66

Product Measures 67
The Proportional-Agreement Statistic 68
The Commonality-Agreement Score 74
Inter/Intraclinician Agreement 76
Product Measures: Summary 87
Process Measures 89
Basic Process Statistics 91
Process-Agreement Statistics 94
Cue-to-Statement Relationship Statistics 99
Process Measures: Summary 112
The Informal Product/Process Measures 115
Informal Product/Process Measures: Summary 123
Summary 125

V. SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS . 127

Summary 127
Introduction 127
The Problem 128
Review of the Literature 129
Design and Procedures 130
Analysis Measures and Results 133
Discussion 137
Study of Product 137
Study of Process 143
Biasing Effects 144
Conclusions 148
Training 149
Check List 151
Recommendations 152

Chapter Page

APPENDICES 157

BIBLIOGRAPHY 262

LIST OF TABLES

Table | Page

1. Correlation (r_{xy}) and Porter Index for Clinic and Home Conversion of Written Diagnosis to RDCL on Four Cases 56

2. Means and Standard Deviations for Subjects Based on Correlation (r_{xy}) for Use of RDCL 58

3. Means and Standard Deviations for Cases Based on Correlation (r_{xy}) for Use of RDCL 58

4. Proportion of Clinician Agreement on Most Frequently Collected Cues, Case 1 69

5. Percentages for Proportion of Agreement on Cues Most Frequently Requested for Four Cases 71

6. Diagnostic Statements Most Frequently Selected From a Standard Check List 72

7. Percentages in Proportion of Agreement for Most Frequently Mentioned Diagnostic Statements on Four Cases 73

8. Commonality Agreement on Cues Collected and Diagnostic Statements Made in Six Diagnoses on Each of Four Cases 75

9. Means and Standard Deviations for Inter/Intraclinician Agreement (Phi and Porter) on Cues for Four Cases . . . 79

10. Means and Standard Deviations for Inter/Intraclinician Agreement (Phi and Porter) on Diagnostic Statements for Four Cases 79

11. Means and Standard Deviations for Intraclinician Agreement for Cues Collected on Four Cases 80

12. Intraclinician Agreement for Two Conversions (Clinic and Home) of Written Diagnoses to the RDCL 82

13. Means and Standard Deviations for Intraclinician Agreement for Diagnostic Statements on Four Cases . . . 83

Table		Page

14. Agreement Statistics on _Cues_ Collected for Four Simulated Reading Cases 86

15. Agreement Statistics for _Diagnostic Statements_ Made on Four Simulated Reading Cases 87

16. Basic Process Statistics Data for Four Simulated Cases in Reading 92

17. Means and Standard Deviations for Value of _r_ on _Cue_ Times, Including Missing Data, for Four Cases 96

18. Means and Standard Deviations for Value of _r_ on _Diagnostic Statement_ Times, Including Missing Data, for Four Cases 97

19. Inter/Intraclinician Phi on Four Time Partitions for _Diagnostic Statements_ and Four Cases 99

20. Average of Percentage of Number of Times Each Relationship Appeared for _Cues_--Case 1 103

21. Average of Percentage of Number of Times Relationships Appeared for _Cues_ Using Hypothesis-Directed and Cue-Directed Inquiry Responses Across All Subjects and All Cases 106

22. Means and Standard Deviations Across All Subjects and Cases for Average of Percentage of Number of Times Relationships for Hypothesis-Directed and Cue-Directed Inquiry Responses Appeared for _Cues_ 107

23. Average of Percentage of Number of Times Each Relationship Appeared for _Diagnostic Statements_--Case 1 109

24. Average of Percentage of Number of Times Relationships Appeared for Diagnostic Statements Using Hypothesis-Directed and Cue-Directed Inquiry Responses Across All Subjects and All Cases 110

25. Means and Standard Deviations Across All Subjects and Cases for Average of Percentage of Number of Times Relationships for Hypothesis-Directed and Cue-Directed Inquiry Responses Appeared for _Diagnostic Statements_ 111

26. Sample 1 of Responses to Informal Questions Relating to Memory (1) and Strategy (2)--Subject 104 118

Table		Page

27. Sample 2 of Responses to Informal Questions Relating to Memory (1) and Strategy (2)--Subject 106 119

28. Sample of Responses to Four Questions Regarding Specific Aspects of Reading--Subject 104, Run 3, Case 4 120

29. Results of Responses by Eight Clinicians to Specific Questions Regarding Four Areas of Reading for Four Cases 122

30. Agreement Statistics on Cues Collected and Diagnostic Statements Made by Eight Clinicians in Each of Two Studies Using Four Simulated Reading Cases 139

A1. Background Information on Reading Clinician Subjects . . 160

D1. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 101, Case 3 . . . 202

D2. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 102, Case 2 . . . 203

D3. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 103, Case 1 . . . 204

D4. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 105, Case 3 . . . 205

D5. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 107, Case 1 . . . 206

D6. Subject Responses to Informal Questions Regarding Memory (1) and Strategy (2)--Subject 108, Case 2 . . . 207

D7. Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 103, Case 1 . . . 208

D8. Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 107, Case 1 . . . 209

D9. Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 102, Case 2 . . . 211

D10. Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 108, Case 2 . . . 212

D11. Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 101, Case 3 . . . 214

Table		Page

D12.	Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 105, Case 3 . . .	216
D13.	Subject Responses to Informal Questions Regarding Specific Aspects of Reading--Subject 106, Case 4 . . .	217
G1.	Procedures for Two Studies of Diagnostic Decision Making Using Simulated Cases and Eight Clinicians . .	228
H1.	Proportion of Agreement on Cues, Case 2	230
H2.	Proportion of Agreement on Cues, Case 3	231
H3.	Proportion of Agreement on Cues, Case 4	232
H4.	Percentages in Total Proportion of Agreement (P.A.) on Cues for Four Cases	233
H5.	Percentages in Total Proportion of Agreement (P.A.) on Diagnostic Statements for Four Cases	234
H6.	Commonality of Agreement on Cues Collected in Six Diagnoses on Each of Four Simulated Reading Cases . .	235
H7.	Commonality of Agreement on Diagnostic Statements Made in Six Diagnoses on Each of Four Simulated Reading Cases	236
H8.	Inter/Intraclinician Agreement (Phi and Porter) on Cues Collected on Four Simulated Reading Cases	237
H9.	Inter/Intraclinician Agreement (Phi and Porter) on Diagnostic Statements Made on Four Simulated Reading Cases	238
H10.	Correlation Coefficient for Cue Times, Including Missing Data, for Case 1	239
H11.	Correlation Coefficient for Cue Times, Including Missing Data, for Case 2	239
H12.	Correlation Coefficient for Cue Times, Including Missing Data, for Case 3	240
H13.	Correlation Coefficient for Cue Times, Including Missing Data, for Case 4	240
H14.	Correlation Coefficient for Diagnostic Statement Times, Including Missing Data, for Case 1	241

Table Page

H15. Correlation Coefficient for Diagnostic Statement Times, Including Missing Data, for Case 2 241

H16. Correlation Coefficient for Diagnostic Statement Times, Including Missing Data, for Case 3 242

H17. Correlation Coefficient for Diagnostic Statement Times, Including Missing Data, for Case 4 242

H18. Average of Percentage of Number of Times Each Relationship Appeared for Cues--Case 2 243

H19. Average of Percentage of Number of Times Each Relationship Appeared for Cues--Case 3 244

H20. Average of Percentage of Number of Times Each Relationship Appeared for Cues--Case 4 245

H21. Average of Percentage of Number of Times Each Relationship Appeared for Diagnostic Statements--Case 2 246

H22. Average of Percentage of Number of Times Each Relationship Appeared for Diagnostic Statements--Case 3 247

H23. Average of Percentage of Number of Times Each Relationship Appeared for Diagnostic Statements--Case 4 248

H24. Individual Case Data (Mean and Standard Deviation) on Porter Statistic, Including Number of Cues and Diagnostic Statements in Domains, for Two Studies . . 249

LIST OF FIGURES

Figure		Page
1.	The Clinical Encounter	7
2.	The Clinical Interaction	7
3.	The Clinical Case	9
4.	The Clinician	11
5.	Random Assignment of Subjects to Clinical Sessions	53
6.	Research Design	54
B1.	Diagram of Clinical Setting	163
I1.	Contingency Table for Calculation of Phi--Example 1	254
I2.	Contingency Table for Calculation of Phi--Example 2	254
I3.	An Example of a Completed Contingency Table	255
I4.	Contingency Table for Calculation of Intra Phi--Example 1	257
I5.	Contingency Table for Calculation of Intra Phi--Example 2	258
I6.	An Example of a Completed Contingency Table	259

LIST OF APPENDICES

Appendix		Page
A.	CLINICIAN BACKGROUND FORM AND INFORMATION	158
B.	CLINICAL SETTING AND INSTRUCTIONS	162
C.	CLINICAL OBSERVATION INSTRUMENTATION	173
D.	CLINICIAN RESPONSES ON FORMAL AND INFORMAL MEASURES .	187
E.	EXAMINER'S NOTES FORM AND LETTER TO SUBJECTS	219
F.	CUE DOMAIN SAMPLE FOR A CASE	224
G.	PROCEDURES FOR TWO STUDIES: OS '77 and OS '78.3 . . .	227
H.	TABLES	229
I.	OBSERVATIONAL STUDY DATA-ANALYSIS SYSTEM (OSDAS) STATISTICS	250

CHAPTER I

THE PROBLEM

Introduction

Developments in the field of reading in recent years have called for more effective and efficient diagnostic techniques and remediation practices (Chall, 1978). Reading researchers have responded to this challenge with a continuing interest in studies of how reading diagnosticians should behave or perform. However, research slights the problem of how reading diagnosticians <u>think about</u> their students' reading problems (Shulman & Elstein, 1975).

Assuming that correct diagnosis of reading problems is important for effective remediation, diagnostic acumen is highly desirable for reading diagnosticians. It might then follow that one approach toward improving the success of therapy or remediation in clinical diagnosis of reading problems would be to improve the diagnostic skills and training of reading clinicians (Hoffmeyer & Bader, 1978; Hoffmeyer, 1979).

At this time, there is little documented evidence of how previously trained and currently trained reading diagnosticians think about their cases. Although there is evidence that some clinical diagnosis is effective in terms of remediation that results in improved reading behavior (Spache, 1976), the reasoning skills that characterize "expert" reading clinicians have not been identified.

Effective diagnostic procedures might be determined through the study of the diagnostic decision-making and problem-solving behavior of expert reading clinicians. Additionally, once these clinical reasoning skills are identified, training programs in reading designed to teach more effective clinical skills might be developed. Reading diagnosticians might be taught how to seek answers to pertinent questions and to determine what information is of prime importance and what information need not be dealt with for a given reading problem. Recent studies have shown this to be effective in teaching clinical skills to physicians (DeDombal et al., 1974; Elstein, 1975; Elstein, Shulman et al., 1976).

Until recently, studies of clinical problem solving and decision making have involved research in medicine (DeDombal et al., 1972, 1974; Schwartz, 1973). However, in a thorough review of a number of theoretical models and research methods on thinking, human judgment, and decision making, Shulman and Elstein (1975) determined that those studies of information processing, decision making, policy capturing, and lens model "have rarely been applied to the investigation of educational problems" (p. 32). They then proposed a variety of ways in which the models and methods reviewed might be valuable in education. One major area of their discussion dealt with possibilities of research in diagnosis and remediation of reading difficulties. Shulman and Elstein stated that "as in so many situations involving clinical judgment, the principles governing decision making are typically unclear" (p. 34). They concluded that it is crucial to examine <u>how</u> the problem solver "sizes up the situation, how the problem is formulated, what is judged to be relevant and what irrelevant, which sources of information are considered useful and which of no importance" (p. 37).

To move in the direction proposed by Shulman and Elstein, there is a need to observe and study the decision-making and problem-solving behavior of "expert" reading clinicians so as to distinguish those reasoning skills that characterize them as "experts." Then, as in medicine (DeDombal et al., 1972), attempts might be made to use these "experts" as models, by means of simulation, in teaching effective clinical diagnosis and remediation in reading.

These previous studies in medicine, dealing with the investigation of clinical problem solving (Barrows et al., 1976; Elstein, Shulman, Sprafka et al., 1978; Vinsonhaler, Wagner, & Elstein, 1977) provided the foundation for research on clinical diagnosis in reading. From these earlier studies, three basic principles have directed the investigation of reading clinicians:

> First, the recognition of the practical value of a well understood theoretic base for empirical research on clinical problem solving; second the recognition of the efficacy of the methods developed in medicine to examine problem solving under well-controlled conditions using simulated cases and stimulated recall interviews; third, the recognition of the need for a systematic program of research studies which share a common methodology (Vinsonhaler, 1979a, p. 4).

The present study was undertaken as part of a larger research effort by the Clinical Studies group of the Institute for Research on Teaching at Michigan State University. The goal of the Clinical Studies Project was "to better understand, both theoretically and empirically, the clinical skills involved in diagnosing and remediating reading problems" (Gil, Hoffmeyer et al., 1979, p. 12). Further, it was hoped that this improved understanding would lead to improved instruction and evaluation of reading clinicians and ultimately to more effective and efficient diagnosis of reading problems.

Purpose

The general purpose of this study was to use the Inquiry Theory of Clinical Problem Solving as a theoretic base in determining how eight experienced, highly trained reading clinicians diagnosed specific reading problem cases (process) and what information they used in making diagnostic decisions about specific cases (product or outcome). Specifically, the purpose was to test three basic components of the Inquiry Theory. These components, subsequently explained as corollaries, are (1) the agreement of reading clinicians in collecting data (cues) on a specific case in order to diagnose a reading problem, (2) the agreement of these same clinicians in making diagnostic statements, and (3) reading clinicians' use of hypotheses to direct their diagnostic clinical inquiry.

The general main line questions, based on the three components of the Inquiry Theory, that were addressed in this study through formal measurement are:

1. Do experienced reading clinicians agree on the data they collect for a specific reading case in making a clinical diagnosis?
2. Do experienced reading clinicians agree on the diagnostic statements they make for a specific reading case in making a clinical diagnosis?

3. Do experienced reading clinicians indicate the use of hypotheses to direct their clinical inquiry in diagnosis?

These three questions are restated more specifically for research purposes in Chapter IV.

Additional secondary-level questions that were addressed in the study through informal assessment are:

1. Do experienced reading clinicians agree on what information should be included in a "good" diagnosis?

2. Do experienced reading clinicians agree on how to go about a diagnosis?

3. To experienced reading clinicians employ a schema in making diagnostic determinations about specific skill areas of reading for a particular case?

These questions also are more clearly defined in Chapter IV.

The next section contains a description of the Inquiry Theory of Clinical Problem Solving and its related corollaries.

Theory

Introduction

For the last few years, a team of researchers at Michigan State University has been actively involved in studying the behavior of clinical problem solvers and in developing a theory to explain such behavior. The development of a formal theoretical structure that can integrate the numerous concepts and empirical findings on clinical problem solving has become known as the Inquiry Theory of Clinical Problem Solving (Elstein, Shulman, Sprafka et al., 1978).

In understanding the important concepts of the Inquiry Theory, it is necessary first to establish the basic definitions and parameters of the theory. Three major assumptions are presently associated with the Inquiry Theory: (1) the clinical-encounter assumption, (2) the simulated-case assumption, and (3) the simulated-clinician assumption.

The Clinical Encounter

The behavioral domain of the Inquiry Theory is known as the clinical encounter and may be defined as the events that occur as a clinician (e.g., a reading clinician, a teacher, or a physician) attempts to solve a problem in a case (a student, client, patient, or patient record) by making a diagnosis (Dx: What is the problem?) and prescribing a treatment (Rx: What can be done to solve the problem?).

The first assumption of the Inquiry Theory is that the clinical encounter involves (1) a clinician, (2) a case, and (3) the interaction that occurs as the clinician analyzes the information provided by the case, makes a diagnosis, and prescribes remediation or therapy. Three basic components, therefore, comprise the clinical encounter: (1) the clinical interaction, (2) the clinical case, and (3) the clinician (Figure 1).

The clinical interaction.--The clinical interaction, or the reciprocal behavior that occurs between the clinician and the case, is a part of the clinical encounter. Key behaviors that have been observed in the clinical interaction are presented in Figure 2. The direction of interaction is indicated by arrows.

The Inquiry Theory attempts to predict only those aspects of the clinical interaction that may be repeatedly observed, i.e.,

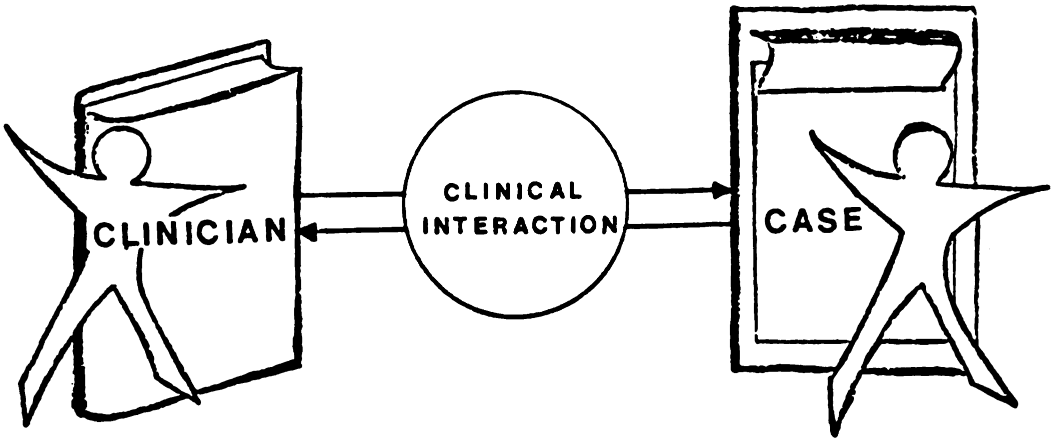

Figure 1.--The clinical encounter.

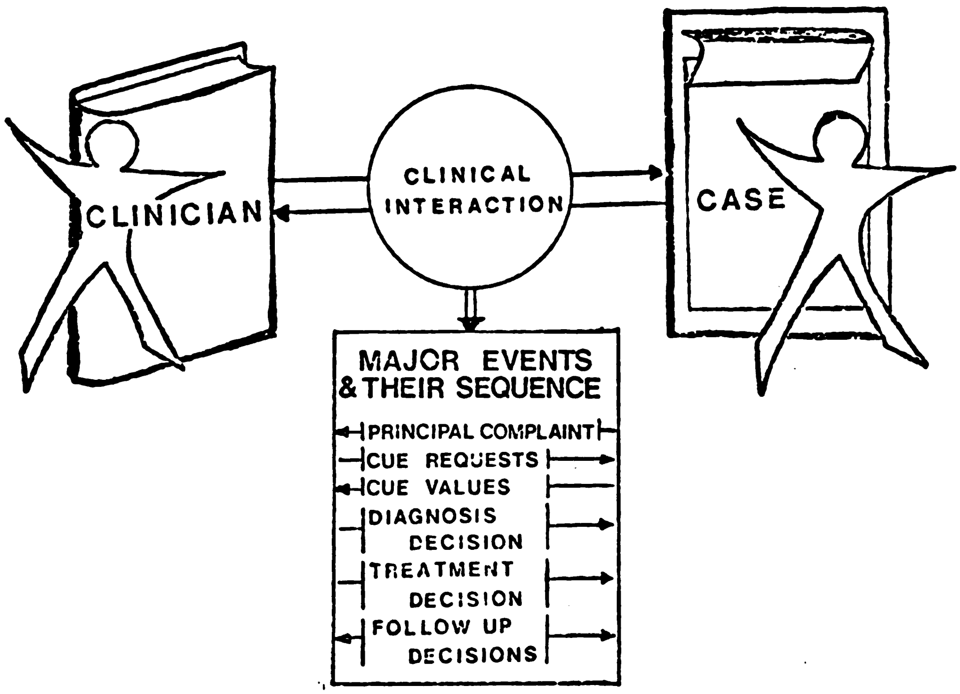

Figure 2.--The clinical interaction.

the features observed when several clinicians interact with the same case or when a single clinician interacts with one case or with several cases (Gil, Hoffmeyer et al., 1979; Vinsonhaler et al., 1977a).

The major elements in the clinical interaction and their sequence include:

1. a principal complaint (statement of a symptom initiated by the case),

2. cue requests (gathering of case information by the clinician),

3. cue values (the clinician's assessment of the significance of the information collected in terms of its relationship to the problem),

4. diagnosis decision (or determination of the problem) by the clinician,

5. treatment decision (selection of treatment seen as being most appropriate), and

6. follow-up decision (determination of treatment efficacy).

<u>The clinical case</u>.--The second component of the Inquiry Theory, the clinical case (Figure 3), involves the second assumption of the theory: "Cases can be effectively simulated (able to elicit some of the same problem solving behaviors [i.e., the clinical interaction] as a real case) by providing the clinician with sets of requested information" (Vinsonhaler, 1978, p. 4).

An alternative to presenting the behavioral domain of a live client to a clinician is to use simulated cases (SIMCASEs). A simulated case is designed to elicit many of the same problem-solving behaviors from the clinician as would a live case. Relevant information (e.g., physical records, background information, test behaviors) can be collected and stored in a file box (manually based SIMCASE) or in a computer file (computer-based SIMCASE) (Lee & Weinshank, 1976). Case simulation is used in the study of the clinical encounter so as to achieve a level of objectivity that is scientifically acceptable and replicable.

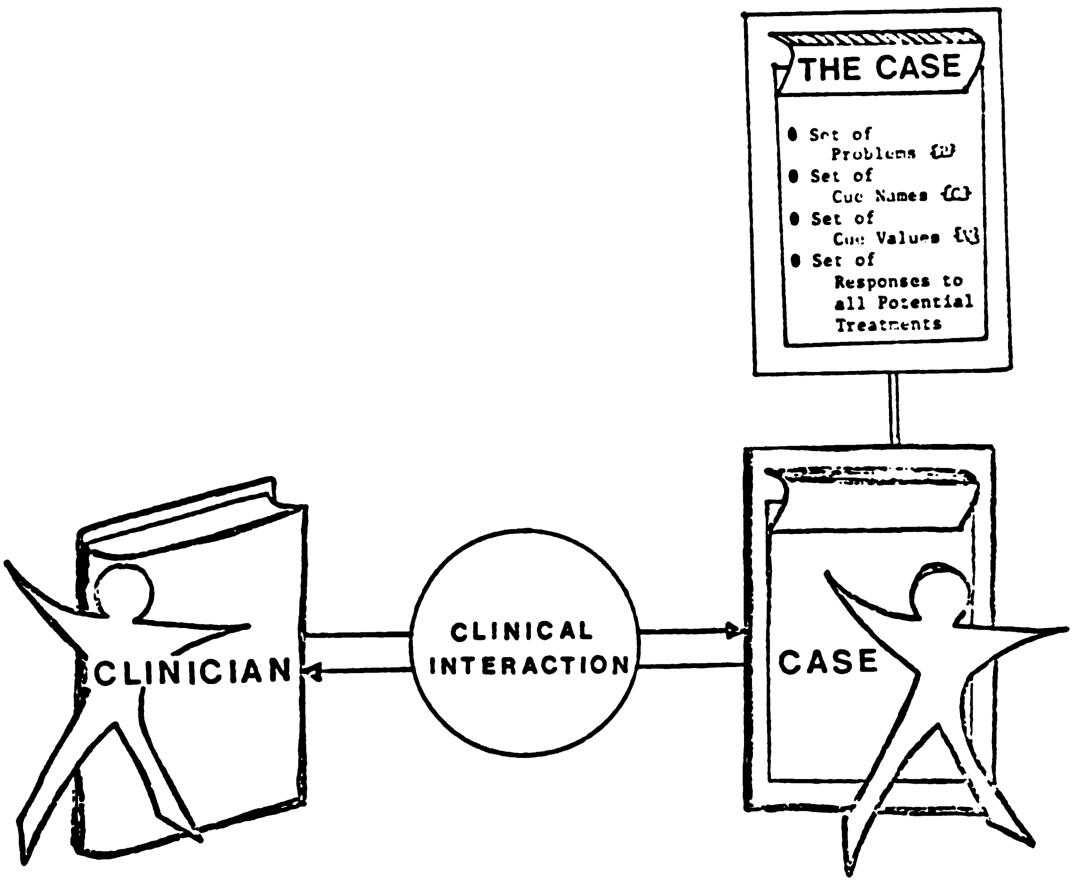

Figure 3.--The clinical case.

A simulated case (SIMCASE) is a set of data representing a client. The cognitive elements of a case include (1) a set of problems (e.g., sight words inadequate), (2) a set of cue names or items of information in a case that might be used to help a clinician make a diagnosis (e.g., Dolch Word List), (2) a set of cue values specifying the client's state with respect to the cue (e.g., 10 percent correct on Dolch Word List), and (4) a set of responses to all the potential treatments (e.g., Work with student on developing a basic sight vocabulary through drills, games, etc., using Dolch Word List).

Although the validity of this assumption in reading has yet to be obtained because of lack of funding and the legal limitations on use of subjects, it has been shown to be valid in medicine (Taylor, Skakun, & Wilson, 1977), where simulated cases have been widely used.

<u>The clinician</u>.--The third Inquiry Theory component, the clinician, concerns the third assumption of the theory and involves those factors that govern the behavior of the clinician in the clinical encounter. This assumption states that "the major events in the clinical interaction are determined probabilistically by the CASE and the CLINICIAN's MEMORY AND STRATEGY" (Vinsonhaler et al., 1977, p. 11).

Research findings in medicine support the simulated clinician assumption (see Elstein, Shulman, Sprafka et al., 1978).

Clinical memory consists of problem, cue values, prescription and treatment descriptions, and the relations between them (Figure 4).

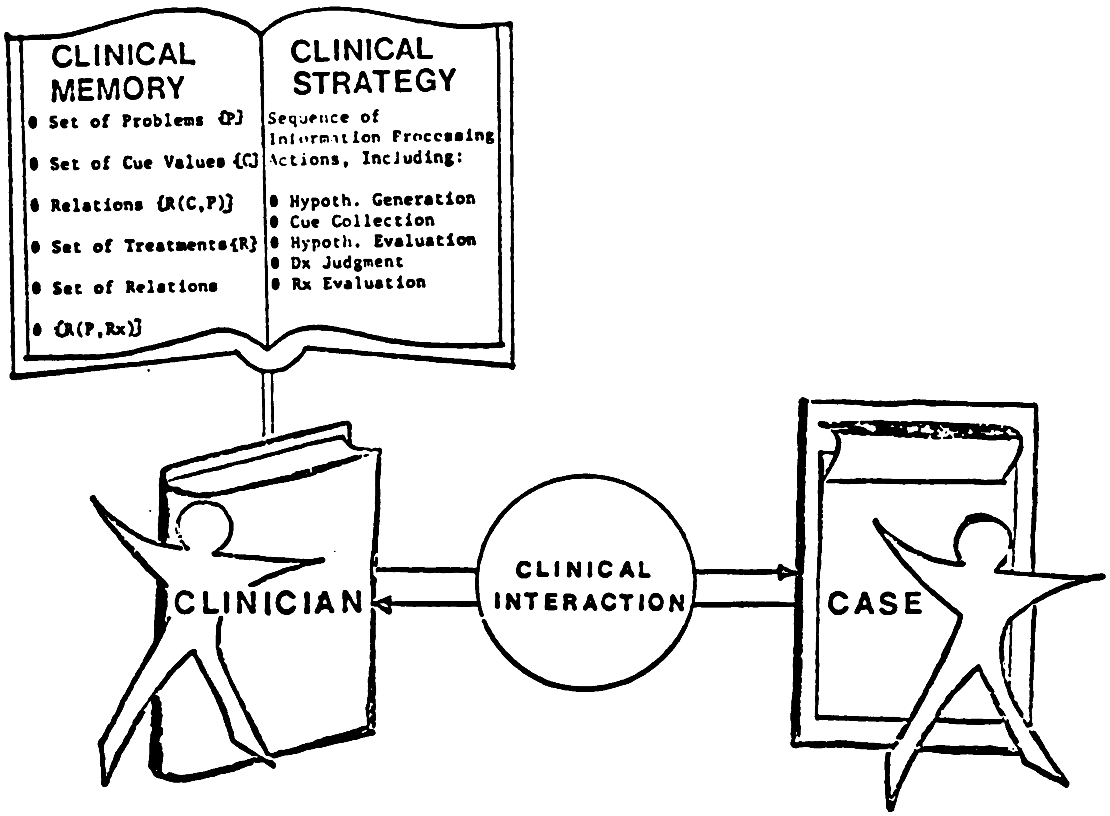

Figure 4.--The clinician.

A clinical memory example from medicine might be as follows:

Problem Representation for Angina Pectoris

Elevated blood pressure
Chest pain = 1.0
Past history ischemia = 1.00
ECG shows RS-T deviations or T-wave inversions = 1.0

Cue Representation for Angina

Elevated blood pressure <40 138/80
<20 120/70

Chest pain
Does the patient report:
(1) chest pain after exercising, e.g., walking up stairs?
(2) chest pain after a heavy meal?
(3) pain localized behind the sternum or radiating to the left shoulder and arm?

Treatment for Angina Pectoris

Restricted exercise
No smoking
Alcohol in moderation
Fat-restricted diet
Nitroglycerin

Prescription for Angina Pectoris

Nitroglycerin (Glyceryl Trinitrate)
Sublingual tablets (.3 to .6 mg) q. 2 h. to q. 3 h. as required

In addition to problem, cue values, and prescription and treatment descriptions, clinical memory consists of a number of important relations:

1. a set of relations between cues and problems used to infer the presence of problems in a given case, based on the cues already collected--R(C,P);

2. a set of relations between problems and cues used to determine which cues should be collected next, in order to confirm or disconfirm the hypotheses currently under consideration--R(P,C);

3. a set of relations between problems and treatments used to evaluate and select treatment plans for a given case and diagnosis--R(P,T); and

4. a set of relations between treatments and prescriptions used to define the specifics of case management for a particular case and treatment plan--R(T,Rx) (Vinsonhaler et al., 1977a).

Clinical strategy consists of a sequence of tasks that translate memory into action. These tasks mainly involve information gathering and information processing as the clinician makes decisions about diagnosis and treatment. Those tasks that were empirically derived from studies of clinical problem solving with simulated cases (Elstein et al., 1978) include the following:

1. Cue acquisition--the process by which the clinician decides which information (cue) should be collected in a "medical history, physical examination, and laboratory work-up" and the relative value of those cues selected. Cues may be chosen on the basis of (a) confirming or disconfirming one or more competing hypotheses concerning the patient's problem or (b) according to some information-gathering routine work-up.

2. Hypothesis generation--the process of retrieving from memory a number of problem formulations (hypotheses) based on (1) some limited number of cues and (2) the relations between the cues and problems R(C,P) that are part of the clinician's memory. Early generation of hypotheses may be used to direct the work-up.

3. Cue interpretation--the process by which case information (CUES) is "evaluated in terms of [its] 'fit' to specific hypotheses."

4. Hypothesis evaluation and diagnosis judgment--the process by which an estimate of the likelihood of each hypothesis being considered is determined. This is done by (1) eliminating unlikely

hypotheses and (2) accepting as the diagnosis those hypotheses with sufficiently high likelihood. "Likelihood is calculated on the basis of the relations between problems and cues R(P,C)" (Vinsonhaler et al., 1977a, p. 12).

5. Treatment evaluation--the process by which an estimate is made of the expected gain from each available treatment for the diagnosed problem. Expected gain is calculated on the basis of the relations between the problem and available treatment plans R(P,T). These relations may include effications, cautions, contraindications, cost, and preference appropriate to the patient or case. A treatment plan is selected on the basis of highest expected gain.

6. Prescription selection--the process by which relations between treatments and prescriptions R(T,Rx) are used to write out the specifics of the case management.

With the knowledge of effective clinical problem solving, computer programs can be developed to simulate a clinician engaged in clinical diagnosis. The memory and strategy of the computer arrive at a diagnosis of a problem. Thus the computer applies a diagnostic process used by human clinicians. One computer simulation system is discussed in the following section.

The Basic Management Information System (BMIS)

The Basic Management Information System (BMIS) computer program makes it possible to simulate a clinician. First, a simulated memory must be described as in the previous medical example of angina. Then a description of a strategy or sequence of information-processing actions (see Figure 4) is presented. Finally, a case is presented and the computer behaves according to the described memory and strategy. Simulated clinicians (SIMCLINs) are used to test and improve the Inquiry Theory by establishing valid deductions and quantitative predictions and comparing them with the clinical problem-solving behavior of real clinicians (Vinsonhaler et al., 1977b).

As will be shown later, the present study was intended to examine certain predictions derived from the existing Inquiry Theory.

Theoretic Implications

This research adhered to an IRT Clinical Studies Project objective of developing and empirically testing clinical problem-solving theory and its application to reading. The present conceptual replication study was therefore concerned with testing theoretic implications of three basic components of the Inquiry Theory of Clinical Problem Solving.

The theoretic implications that seem relevant to this study are (1) the effect of clinical memory and strategy on cue-collection agreement, (2) the effect of clinical memory and strategy on the diagnostic (Dx) agreement of clinicians, and (3) the effect of clinical memory and strategy on hypothesis generation.

These three components may be restated as corollaries or propositions derived from the informal* Inquiry Theory and implications noted for each of the three.

---

*The Inquiry Theory is often divided into three parts: (1) the formal theory, a set of Fortran computer programs--BMIS; (2) the interpreted Inquiry Theory, simulation study results based

1. Cue (Cx) Agreement Corollary

Informal statement: The greater the number of common cues represented in memory, the greater the number of common cue elements in the clinical interaction.

Formal statement: If $N\ (Ci \cap Cj) \geq N\ (Ck \cap Cl)$, then $N\ (Cx_i \cap Cx_j) \geq N\ (Ck_k \cap Cx_l)$, all else equal, where Ci denotes the cue component of clinical memory and Cx_i denotes the set of cues present in the i^{th} clinical encounter.

Implications of cue (Cx) agreement corollary: Following arguments similar to those presented for the Dx agreement corollary, two predictions may be offered for observational study results. Those predictions follow.

a. Same clinician, same case versus different clinicians, same case: Assuming that the cognitively formed cue memory is stable (the common or shared elements do not vary or change significantly over time), the agreement in cue selection by the same individual should be equal to or greater than the agreement between individuals. Since $N\ (Ci \cap Ci) \geq N\ (Ci \cap Cj)$, then $N\ (Cx_i \cap Cx_i') \geq N\ (Cx_i \cap Cx_j')$, assuming all other factors are constant.

b. Commonality of cue selection: Given a set of cues, if common elements exist in the clinical memories, then there should be "commonality" or agreement among clinicians on cues selected for any given case.

---

on BMIS program data; and (3) the informal theory, a natural English-language summary and interpretation of simulation studies.

<u>Measurement problems</u>: Many of the problems associated with the use of numerocity of sets in diagnosis (Dx) predictions are also common to cue (Cx) predictions. Therefore, both proportional frequency and correlational measures of agreement are used for measuring cue "commonality."

2. <u>Diagnostic (Dx) Agreement Corollary</u>

<u>Informal statement</u>: Given any two diagnoses (statements concerning the problems or conditions of the client) using the same techniques and based upon the same case, the greater the number of common or shared elements in clinical memory (problems, cues, and the relations among them), the greater the number of common or shared elements in the diagnoses, assuming all other factors are constant.

<u>Formal statement</u>: If $N\ (Pi \cap Pj) \geq N\ (Pk \cap Pl)$, then $N\ (Dx_i \cap Dx_j) \geq N\ (Dx_k \cap Dx_l)$, where Pi denotes the set of state or problem descriptions, i.e., strengths and weaknesses of the case, and Dx_j denotes the set of diagnostic statements about the case.

<u>Implications of diagnostic (Dx) agreement corollary</u>: Based on the Observational Study, 1977 (see Chapter III) data analysis to date, there seem to be two implications for the present research study. Those implications follow.

a. <u>Same clinician, same case versus different clinicians, same case</u>:

<u>Informal statement</u>: Assuming that the cognitively formed diagnosis (Dx) is accurately reflected in a written diagnosis and that the clinical memory remains reasonably stable (the common or shared elements do not vary or change significantly) over short time

periods, then the number of common elements between diagnoses prepared by the same clinician on two forms of the same case should be greater than or equal to diagnoses prepared by different clinicians on the same case or on alternate forms of the same case.

Formal statement: If $N\ (Pi \cap Pi) \geq N\ (Pi \cap Pj)$ for $i \neq j$, then $N\ (Dx_i \cap Dx_i') \geq N\ (Dx_i \cap Dx_j)$.

b. Commonality of diagnostic statements:

Informal statement: Given a set of diagnoses, if common elements exist in the clinical memories, then two or more clinicians should have common or shared problem or state descriptions in their diagnoses; i.e., there should be a commonality of diagnostic statements.

Formal statement: If $N\ (Pi \cap Pj \cap Pk\ .\ .\ .\ \cap Pn) \neq 0$, then $N\ (Dx_i \cap Dx_j \cap Dx_k\ .\ .\ .\ \cap Dx_n) \neq 0$. Hence an index of commonality should yield nonminimum values; i.e.,

$$I\ (\text{Dx statement } k) = \frac{\text{Number of Dx's including statement}}{\text{Number of diagnoses}} > \frac{1}{N}$$

Measurement problems: The numerosity of the sets of common diagnostic elements or statements $N\ (Dx_i \cap Dx_j)$ seems to be an inadequate basis for measurement because the number of common elements in the diagnoses depends on the number of diagnostic categories (background information, Dolch Word List, etc.) used in the set (Dx_i and Dx_j); i.e., the larger the set, the more likely one is to get agreement. Rather, various "normalized" statistics have to be used to evaluate the various hypotheses stated previously. These include a diagnostic (Dx) commonality score and various types of correlation and frequency matches.

3. Hypothesis (Hx) Generation Corollary

Informal statement: Clinical strategy based on deductive reasoning, using hypotheses to direct inquiry, should show a tendency toward early hypothesis and observation generation. Hence, the clinical interaction of a deductive reasoner may be characterized as follows: The first statement of hypothesized or observed states of the clinician (later included in the diagnoses or otherwise dropped) generally occurs in the first half of the interaction.

Formal statement:

$$Xsg = \sum_{I=1}^{4} (I \cdot P_{SI}), \text{ where } P_{SI} = \frac{N \text{ (new statements } I^{th} \text{ quarter)}}{N \text{ (total new statements)}}$$

The hypothesis statement generation score is equal to the value in the first, second, third, and fourth quarters of the sum of the product of I (the quarter = 1,2,3,4) and P_{SI} (the proportion of original, not previously stated, descriptions of the client's clinical states occurring in the I^{th} quarter). In general, deductive reasoners are characterized by low Xsg scores, e.g., less than 2.0, whereas inductive reasoners are characterized by high Xsg scores, e.g., greater than 2.0.

Implications of hypothesis (Hx) generation corollary: The major implication of the Hx corollary is that clinicians having hypothesis statement generation (Xsg) scores above 2.0 (hypotheses generated in the third and fourth quarters of the session) are probably not using deductive reasoning. Rather, such individuals must be using some approach that directs their inquiry on the basis of something other than hypotheses. Given recent research findings with medical and reading clinicians, another pattern of clinical strategy has been postulated--that of inductive reasoning. Inductive reasoners may be defined as clinicians who collect cues on a presently unknown basis, perhaps largely at random from some fixed preferential set, and who then interpret those cues by generating hypotheses. It then appears that once these clinicians are satisfied that a sufficient number of cues has been collected, they then attempt to state a diagnosis (Vinsonhaler, 1979a).

Research involving the hypothetico-deductive approach or early generation of hypotheses in clincial diagnostic inquiry has been investigated in medical studies involving the Inquiry Theory (see Elstein, Shulman, Sprafka et al., 1978).

For the sake of clarification and to avoid confusion over the terms deductive and inductive as used in the psychological literature, discussions of hypothesis generation in this study refer to deductive reasoning as hypothesis-directed inquiry and to inductive reasoning as cue-directed inquiry. (See Process Measures, Chapter IV.)

Summary

The purpose of Chapter I has been to direct attention toward the need for the investigation of clinical problem solving in reading. The need is threefold: (1) there is practical value in using a theoretical base for empirical research, (2) there is efficacy in using scientifically based methods (medical studies) to examine clinical problem solving in reading using simulated cases, and (3) there is augmentative value for research on clinical problem solving when research studies share a common methodology (Vinsonhaler, 1979a).

The purpose of this study was to investigate three components of the Inquiry Theory of Clinical Problem Solving, which were described in the second part of Chapter I. The Inquiry Theory postulates that the clinical encounter involves a case and a clinician and is characterized by the interaction that occurs between the case and the clinician's memory and strategy. The three major questions addressed in this study and that relate to the Inquiry Theory are: (1) Do experienced reading clinicians agree on the data they collect for a specific reading case in making a clinical diagnosis? (2) Do experienced reading clinicians agree on the diagnostic statements they make for a specific reading case in making a clinical diagnosis? and (3) Do experienced reading clinicians indicate the use of hypotheses to direct their clinical inquiry in diagnosis?

Overview

This dissertation is a summary of the application of the Inquiry Theory of Clinical Problem Solving as it relates to the clinical decisions of reading clinicians. In Chapter II, the literature pertinent to clinical diagnosis is reviewed. It includes a discussion of research procedures in clinical problem solving, problems inherent in reading diagnosis, and closely related research involving the

Inquiry Theory. In Chapter III, the design for the study is explained. This explanation includes a description of the sample, the operational procedures, operational and reliability measures, research questions, and the design.

The analysis methods and results are presented in Chapter IV. The chapter format deviates from that which is often used in the presentation of analyses, the reason being the somewhat complex nature of the methods of analysis, many of which were calculated on data generated by computer programs developed specifically for this and related studies. In Chapter IV, one analysis measure is explained, and the data for that measure are presented before the next measure and its corresponding data are presented. An attempt is made to relate each measure to the Inquiry Theory.

Turning now to a review of literature related to clinical decision making in reading diagnosis, the focus is on what ideas pervade reading professionals' discussion of diagnosis. Additionally, it is shown that the pervasive trend is changing, as evidenced by new theoretically based research in the area of clinical diagnosis.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

In the reading literature there is no definition or general consensus about what constitutes a "good" diagnosis. Neither is it obvious what procedures a clinician should follow in reaching an effective and efficient diagnosis, one that will direct equally effective and efficient remediation. Additionally, there is disagreement in the reading field over terms used to describe reading problems. For example, the word "dyslexia" has become so nebulous that it is virtually useless; the same is true of such terms as "perceptual deficit" and "minimal brain damage." The literature is replete with examples. Compounding the problem is the debate among reading specialists over causation. These and other dilemmas in the reading field have undoubtedly contributed to the dearth of systematic, scientific, and theory-based research in an area most vital to successful reading diagnosis and remediation--that of clinical decision making.

In the psychological literature, related research can be variously represented on cognition, thinking, human judgment, problem solving, and decision making, with terms referring to different research paradigms and models. Although such terms are conceptually unclear, their common relationship in realistic task environments seems to be accepted (Shulman & Elstein, 1975).

The manner in which processed information is acquired seems to be as varied as the terms used to describe it. A review of some of the approaches used to study the intellectual process of subjects as they render judgments, solve problems, and make decisions provides the impetus for ways in which these research approaches can be extended to the study of relevant issues and concerns in reading as well as in other areas of education.

In Chapter II, a review of literature relative to clinical decision making in reading and to the questions being addressed in this study is presented. The review is divided into three major parts: (1) Research Procedures in Clinical Problem Solving, (2) Problems Inherent in Reading Diagnosis, and (3) The Inquiry Theory Research.

Research Procedures in Problem Solving

For purposes of this study, the discussion of research procedures in problem solving is confined to three methods: total-task or process-tracing studies, in-basket techniques, and tab-item methods.

Total-Task Studies

One means of studying problem solving involves describing the intellectual processes of subjects as they make decisions and render judgments. This "process-tracing" approach typically characterizes human thinking and problem solving through verbal reporting or restatement in a computer-simulated program.

Process-tracing studies may involve partial or subtask examination or may investigate a total task. "Total task studies investigate the sequential character of information seeking that leads to judgments or decisions. They use forms of simulation to represent the task environment. . ." (Shulman & Elstein, 1975, p. 5).

One example of a total-task, high-fidelity study involved the problem solving of physicians (Elstein, Kagan, Shulman, Jason, & Loupe, 1972; Sprafka & Elstein, 1974). Elstein and his colleagues attempted to describe the cognitive processes of physicians beginning with an initial patient encounter and continuing through the final diagnosis. Thinking-aloud protocols of physicians and simulated patients were analyzed using data, hypotheses, and the relation between them. The general model of medical inquiry that emerged included four major activities: (1) acquisition of data or cues, (2) generation of hypotheses, (3) interpretation of data or cues, and (4) evaluation of hypotheses. The most universal characteristic of the sample of physicians and problems in the study was the early generation of hypotheses. Diagnostic accuracy was associated both with slightly higher thoroughness of acquisition of data and with greater accuracy of interpretation of data. There was no correlation between accuracy of interpretation and thoroughness of cue acquisition.

Research on clinical problem solving using verbal reporting as a measure of process tracing has been studied in a variety of ways: Clarkson (1962)--process tracing to model the decision process of a bank trust investment officer. De Groot (1966)--investigation of the thought processes of chess players. Kleinmuntz (1968)--diagnostic problem solving among clinical psychologists interpreting MMPI profiles and clinical neurologists using simulated data in their specialty.

Discussing his own work, De Groot (1966) included a set of principles that, according to Shulman and Elstein (1975), "serve as a credo for the process-tracing approach, ethnographic style:"

> First, the research is directed toward systematic description of cognitive phenomena rather than to strict hypotheses testing. Second, we keep machine simulation in mind, but we hardly do it as yet. Third, the experimental settings are often more like real-life than the strictly controlled artificial conditions of the laboratory. Fourth, extensive use is made of introspective techniques of various kinds. Fifth, as a result, protocol coding and interpretation are of crucial importance (and consume a large part of our time). Sixth, prospective outcomes are expected to be primarily valuable to the extent we succeed in providing adequate, systematic process descriptions, possibly to be used as a basis for simulation (pp. 19-20).

In-Basket Studies

Another approach to the study of decision making is the in-basket technique. The decision maker in the in-basket studies receives the inputs for decisions he must make. The in-basket, after which the technique was named, might contain letters, reports, or messages that would need an action or a decision.

Shulman (1965) and Shulman, Loupe, and Piper (1968) created a modification of the in-basket approach of representing decision making.

> . . . to simulate aspects of a classroom's problems. This made it possible to study teacher inquiry behavior under circumstances

> in which subjects would function in a highly unstructured problem-rich task environment. Rather than focusing on teacher behavior as such, the variables of interest were problem sensitivity, use of diverse information sources, use of time (tasks had no time limit), quality of decisions, task organization, sequence of activities, and the like (Shulman & Elstein, 1975, p. 8).

Using in-basket-type techniques, Hemphill, Griffiths, and Frederiksen (1962) studied the behavior of educational administrators.

Tab-Item Methods

Tab-item methods of studying problem solving increase the objectivity and reliability of interpretation and reduce problems of coding, analysis, and interpretation by predesignating the available items of information or choices of action available to the subject. There is no introspection or thinking aloud.

In a study of troubleshooting performance, Glaser, Damrin, and Gardner (1954) conducted the earliest published tab-item study. In that study, the performance failure of a piece of electronic equipment was described, and a list of all possible tests a troubleshooter might make in order to locate the problem was given. A paper tab covered the test-results information. By removing the tab covering, the subject would leave a record of the steps he had taken in determining the source of difficulty.

Rimoldi (1955, 1961) and Rimoldi, Devine, and Haley (1961) were responsible for the development of a large number of sequential problem-solving tests using tab-item methods. They experimented with a variety of scoring procedures that compared the subjects' information-gathering sequence either logically defined or defined by criterion-group performance.

Problems Inherent in Reading Diagnosis

Before understanding what might be involved in the process of reading diagnosis, it seems judicious to note some problems that may be considered inherent in the diagnostic process. Although not all-inclusive, three interrelated problem areas relevant to this study are (1) the essence of reading diagnosis, (2) the debate over causation, and (3) the lack of standardized terms.

The Essence of Reading Diagnosis

The word "diagnosis" is derived from two Greek roots, "dia," meaning thorough or thoroughly, and "gnosis," meaning knowledge. The literal meaning of "diagnosis" is a thorough knowledge, whereas its medical meaning is the determination, by examination, of the nature and circumstances of a diseased condition. The definition of the word "diagnosis" that seems most applicable to the study of reading problems is "a determining or analysis of the cause or nature of a problem or situation" (Random House Dictionary of the English Language, 1966). If one is to take this definition literally, one would be more concerned, it seems, with the etiology of reading problems. It is not surprising to find this emphasis on causes, since the word "diagnosis" has long been associated with the medical profession, and medical terminology tends to label learning problems in terms of causation or etiology.

The word "remedy" or "remediation" has to do with a "healing treatment" or restoring to a "natural or proper condition," i.e., a therapeutic concern (Random House Dictionary of the English Language, 1966).

Although diagnosis and remediation might be defined separately, they are probably most frequently considered part of the same process, occurring simultaneously during treatment (Spache, 1976). However, some in the reading field would tend to dichotomize the diagnostic process. For example, Bond and Tinker (1957) described two types of diagnosis--etiological and therapeutic. Etiological diagnosis is concerned with causation and why the child is in difficulty. According to Bond and Tinker, to prevent reading problems, it would be useful to know that a child had been absent for a month in first grade, but such knowledge would not be useful "for the immediate job of correcting a reading disability that began several years earlier" (p. 127).

Harris and Sipay (1975) also discussed the limited usefulness of etiological diagnosis; they concluded: "Time spent on attempting to determine etiology can often be more profitably spent on helping children to overcome their present problems" (p. 242). They were careful to point out, however, that etiology is important from the standpoint of prevention and correction, but they also felt that opinions about causation are "unproved hypotheses" and that more research is needed in this area.

Continuing the idea of a dichotomized diagnostic process, in a therapeutic diagnosis the concern is with present conditions and situations for the child in order to give direction to reading instruction. Knowing current strengths and weaknesses is more important in therapeutic diagnosis than is awareness of a temporary hearing impairment that occurred several years ago (Bond & Tinker, 1957).

Part of the process of reading diagnosis is collecting information on why a student is having difficulty, and this necessarily involves giving tests. However, diagnosis is not testing. Rather, it is an intelligent interpretation of information based on theoretical knowledge and practical experience. It involves knowing (1) what questions to ask, i.e., what information to obtain that will aid in comprehending the reading problem; (2) how to interpret correctly the meanings of information; and (3) how to understand the interrelationships of information and meanings.

Another area of diagnosis that seems to need clarification is that of the actual diagnostic procedures themselves. In other words, should one follow certain steps in making a reading diagnosis? Although the answer is again anything but definite, a look at the literature shows that several experts in the reading field have supported various sequential procedures that reading diagnosticians should follow. Others have pointed out factors to be considered in a diagnosis.

After the collection of diagnostic evidence, Harris and Sipay (1975) suggested the following procedures:

1. consider complete picture and arrive at conclusions about what child's major difficulties in reading are,
2. determine most reasonable explanations of how these difficulties have come about,

3. decide what persisting handicaps may impede progress,
4. determine what remedial procedures should be employed to overcome handicaps,
5. use periodic checks of remedial work with formal and standardized tests to determine effectiveness of procedures,
6. alter remediation if necessary, and
7. retest to consider pupil progress (p. 346).

Robinson (1956) named the following principles of diagnosis in what appears to be a sequential ordering:

1. secure as much information as possible
2. obtain highly accurate level of reading ability
3. administer standardized reading survey test
4. analyze data to determine if there is a reading problem
5. make detailed analysis of the problem
6. identify factors inhibiting reading progress
7. collate all data and interpret results
8. make appropriate recommendations for remedial therapy (pp. 152-53).

Robinson's approach to data collection seems to be that of amassing quantities of materials without specification of importance. On the other hand, Bond and Tinker (1957) emphasized a more efficient approach to data collection for reading diagnosis. They discussed three levels of diagnosis: (1) general diagnosis for all children or those for special study, (2) analytical diagnosis to explore specific strengths and weaknesses and only in cases where warranted, and (3) case-study diagnosis, individual diagnosis in which reading skills and abilities need careful study. At the case-study level, Bond and Tinker (1957) gave what appears to be a sequential procedure for data collection:

1. secure information from standardized reading diagnostic tests or procedures
2. study child for appraisals of his mental capability, vision, hearing, and physical characteristics
3. consider child's reactions to his reading disability
4. evaluation of environmental factors--home, school, community, etc. (p. 131).

Recognizing the differential nature of diagnosis and that "opinions as to causation remain in the realm of unproved hypotheses," Harris and Sipay (1975) advised reading specialists and teachers to follow this sequence in dealing with children who appear to have a reading problem:

1. determine the individual's general level of reading achievement and compare it with his potential; if a reading problem exists,
2. determine the learner's specific reading skill strengths and weaknesses,
3. determine which factors are most probably hampering the child's ability to learn <u>at that time</u>,
4. remove or lessen those factors that can be controlled or corrected, either before or during remedial treatment,
5. select the most efficient and effective way to teach the needed skills,
6. conduct a program of skill mastery, and
7. refer to an appropriate clinic or agency any child who does not respond to treatment after a reasonable period of time (p. 242).

Although he did not discuss the specifics of data collection, Carter (1970) concerned himself with four levels of diagnosis, which he presented on a schematic scale. The four levels are (1) Identification of Problem, (2) Classification, (3) Identification of Reading Needs, and (4) Determination of Causal Factors. At the upper or fourth level, Carter then emphasized use of the following procedures, though not necessarily in sequence:

1. Identify the problem and possible causal factors.
2. Assume and reject hunch after hunch until one can be accepted tentatively.
3. Discover possible determinant and explain consequential relationship.
4. Predict that with treatment the disability will be overcome. The clinician must verify this prediction (p. 20).

Although stressing the continuous nature of diagnosis, Wilson (1977) pointed out that the reading specialist is in a flexible diagnostic role, having three levels of clinical diagnosis from which to choose: (1) initial screening or brief, concise evaluation of student's reading skills (approximately one hour); (2) selective testing (not overtesting), and (3) case study or in-depth testing.

Bond and Tinker (1957) mentioned eight principles or aspects of diagnosis:

1. Direct diagnosis toward methods of improvement.
2. Have the diagnosis go beyond appraisal of reading skills and abilities.
3. Make the diagnosis efficient--don't overdiagnose.
4. Collect only important information by most efficient means.
5. Use standardized test procedures when possible.
6. Use informal testing when diagnosis needs to be expanded.
7. Formulate diagnosis on basis of "patterns of scores."
8. Make diagnosis a continuous process (p. 126).

According to Harris (1961), the term "diagnosis" involves two major components: discovery and exploration. The discovery part of diagnosis involves a careful study to determine the nature of the reading condition, whereas the exploration part concerns the causes of the reading difficulty.

In determining the nature of the reading condition, Harris considered both informal teacher-made tests and standardized tests. The areas of reading that he emphasized were (1) reading level; (2) comprehension; (3) rate, fluency, and accuracy in oral reading; (4) word-recognition skills; and (5) learning potentialities in word recognition.

Harris further believed that making a reading diagnosis involves understanding the dimensions of the student's reading performance and of those factors, both present and past, that contributed to the difficulties manifested. He stated that although tests may provide insights and facts needed, "the heart of diagnosis is not testing" but rather involves

1. intelligent interpretation of facts in the light of theoretical knowledge and practical experience,
2. knowing what questions to ask,
3. knowing how to select procedures, including tests, which can supply needed facts,
4. knowing how to interpret the meaning of findings, and
5. comprehending the interrelationships of these facts and meanings (pp. 220-21).

Spache (1976) characterized diagnosis as:

1. a continuous process of testing, observing and hypothesizing in a flexible trial-conclusion strategy,
2. pragmatic and directly related to remedial practice,
3. eclectic and thorough,
4. a constant exploration of the student's strengths and recognizes variability from one subskill to the next,
5. broad enough to explore all possible causes of problems, and
6. only temporary, supportive help for the student (p. 9).

According to Smith and Dechant (1961), knowing the student's strengths and weaknesses is the essence of reading diagnosis. It is important, they said, to study the student's instructional needs based on the expectancies of his chronological age, mental age, and grade placement in light of his general abilities and reading potential. They stated that identifying causal factors in regard to reading development is also a part of diagnosis.

The way one conducts the search (i.e., gathers information) for factors contributing to a reading problem is most likely influenced by one's beliefs about causation (VanRoekel, in progress).

The Debate Over Causation

For centuries, philosophers have been pondering the meaning of causation. When two types of traits, events, or actions can be observed to occur simultaneously more frequently than could be expected on the basis of chance, and if one is consistently preceded by the other, it might be easy to assume that the first is cause and the second effect. This correlation, or the fact that two or more measurable characteristics tend to be found together, does not prove causation. In fact, simple causal relationships can seldom be established in the study of people. This is true of reading disability, said Harris and Sipay (1975). They stated:

> In a particular child with a reading disability several characteristics may be found, each of which has been shown by research to be somewhat correlated with reading disability. To determine which of these characteristics may have interfered with the child's learning to read, and their relative importance in this case, is a difficult detective job at best and often cannot be solved.

> For this reason, it is safer to discuss the correlates of reading disability than the causes. Correlation can be readily demonstrated or disproved; causation is much more difficult to establish (p. 239).

Another issue that Harris and Sipay cautioned about in drawing conclusions with regard to cause and effect was the idea that strength in one area of reading can compensate for weakness in another area. Research by White and White (1972) and by Bell and Aftanas (1972) showed that good and disabled readers differed primarily in the number of abilities in which they showed special immaturities rather than strengths and weaknesses. Weak abilities, said Harris and Sipay (1975), may combine to have a causal effect that could not be produced by one weakness alone.

The following resolution prepared by the Disabled Reader Committee was approved by the 1972 Delegates Assembly of the International Reading Association:

> There is no single cause for reading disabilities. Reading problems can be caused by a multiplicity of factors all of which are probably interrelated. Just as there is no single etiology, there is no one choice of intervention. For these reasons we deplore the action of those individuals and institutions who suggest that their methods are infallible, appropriate and optimal for every child, and universally efficacious (Harris & Sipay, 1975, p. 241).

The search for contributing or causal factors that one conducts in specific reading disability cases may be strongly influenced by one's background and beliefs about causation of reading problems. Neurologists, for example, have concerned themselves with what they call "congenital word blindness or dyslexia." They see this as a constitutional condition, often of a hereditary nature, and often accompanied by other communication difficulties such as problems in listening, speaking, spelling, handwriting, and written composition (Penn, 1966; Gomez, 1972; Rosenthal, 1973).

According to Harris and Sipay (1975), regarding causation:

> From a practical standpoint, the aim of a thorough diagnosis is not to fix the blame for the child's difficulties, but to discover each of the many conditions that may require correction. A person who develops an enthusiasm for any one theory of causation can frequently find evidence of the handicap he looks for, but is likely to overlook many other significant complications while doing so. An unbiased search is needed for a really comprehensive and satisfactory diagnosis. This usually requires the combined efforts of professionals from several different professions (p. 310).

Recognizing the interrelationships of causal factors in the exploration of reading problems was Harris's (1961) approach. He believed that the purpose of a thorough diagnosis is not to "fix the blame" for the student's reading deficiency, but to determine each of the multiple conditions that might be corrected (i.e., more a remediation or therapeutic approach).

The debate over causation, whether one takes a singular or pluralistic view, or dichotomizes in the etiological or therapeutic sense, is directly related to the issue of terms used to describe a reading problem. Thus, depending on who is describing the educational problem, the student might be called reading disabled, retarded reader, perceptually handicapped, or learning disabled.

Lack of Standardized Terms

Despite the fact that reading problems have been studied for many years, terminology in the reading profession is not yet standardized. Those definitions that do exist are, according to Spache

(1966), little more than "armchair descriptions and lack the pragmatic definitions needed" (p. 22).

One example of the lack of agreement among reading specialists in defining terms is the wide discrepancy in estimates of significant retardation in elementary schools. The estimates vary from 10 or 12 percent up to 30 percent or more. De Boer and Dallman (1960) believed that the differences may be largely due to a lack of agreement about "what constitutes retardation" (p. 267).

Smith and Dechant (1961) defined reading retardation as reading below one's "present general level of development" (p. 420), and including physical, emotional, social, and mental development limits.

Durrell (1940) believed that reading retardation of six months in first grade is more serious than a deficiency of a year or more in the sixth grade (p. 279).

Harris (1953), on the other hand, did not consider that a first-grade student has a reading problem unless his reading age is at least six months lower than his mental age. In grades four and above, Harris defined the problem as a discrepancy of a year or more (p. 299).

Bond and Tinker (1957) grouped disabled readers into four descriptive categories, according to problem severity:

1. Simple retardation includes those students whose reading ability is somewhat immature but balanced.
2. Specific retardation involves children who are low in one or more types of reading but are competent in the basic reading skills.

3. Limiting disability includes those children deficient in basic reading skills, which precludes further growth in reading.

4. Complex disability involves children whose reading growth is inhibited by basic skill deficiencies and who exhibit other accompanying problems such as physical or personality handicaps (pp. 81-82).

Spache (1976) defined the disabled reader as one who

1. is retarded in a number of major reading skills (such as rate, vocabulary, comprehension or word analysis);
2. is retarded by one year or more, if in the primary grades, or by two years or more if older (one year or more at primary level, two years from grades four to eight, and three years or more at secondary should be dividing line for pupils who can be dealt with in the classroom and those who need special clinic help);
3. is an individual who has had normal opportunities for schooling; and
4. has continued to show this degree of retardation below his sociocultural peers despite corrective efforts (pp. 4-8).

In addition, Spache pointed out certain mitigating factors that must be considered in identifying disabled readers. These include the student's sociocultural status, the nature of his reading difficulties, the degree of retardation below a level common to his peers, the duration of his problem, and the need for special professional assistance beyond what has been or can be done in the classroom. Spache said that other factors, although not part of the operational definition of a reading disability, should also be considered before selecting students for intensive diagnosis and remedial treatment. These include estimates of treatment duration in light of the severity of the problem, the student's age and grade placement, his I.Q., and other background information.

It appears, then, that basic disagreement affecting reading diagnosis might stem from the lack of standardized terminology or taxonomy. This has complicated communication both within and between professions (Whitecraft, 1971). It is not uncommon to find the same term used with different meanings or the same condition given various labels. As Brown and Botel (1972) noted, further complications are caused by the confusion of causes with symptoms, diagnostic criteria, and correlated characteristics.

The next area of research being presented represents an attempt by some educators at Michigan State University to lessen the confusion in the field of reading diagnosis. This attempt was made by linking one discipline, that of reading, with another, more scientific and precise discipline, namely medicine. The common bond between them is clinical decision making, but the foundation is provided by the Inquiry Theory of Clinical Decision Making, known as the Inquiry Theory.

The Inquiry Theory: Related Research

Introduction

To comprehend fully the magnitude of the Inquiry Theory, it seems important to review the major historical or sequential events that led to the development of the theory and to present further those subsequent research efforts that are most closely aligned with the present study.

The Medical Inquiry Project

The Medical Inquiry Project is an extensive observational study of clinical data gathering and information processing among expert physicians, which led to the development of the Inquiry Theory of Clinical Problem Solving. One major purpose of the project was to study problem solving and reasoning in a complex task environment where previous experience was clearly relevant and the data inherently probabilistic. Medical practice seemed to provide the best opportunity for studying reasoning under those conditions. The resulting model of medical reasoning was thus derived mainly from the intensive study of a few medical problems worked up by approximately two dozen physicians (Elstein et al., 1972). The pursuit of these studies of medical reasoning led to a computer simulation of medical thinking.

From the observational study of physician performance, three major variables emerged: cues, hypotheses, and the relationship between them. Cues are items of data or information in a case that may be used by the physician to help him make a diagnosis. Hypotheses are possible diagnoses that the physician uses to direct inquiry and rubrics in short-term memory under which the cues or data may be stored. The relationship between cues and hypotheses involves the interpretation of cues and testing of hypotheses.

Clinical tasks are actions directed toward alleviating problems. There are two types of clinical models, the diagnostic and the therapeutic. In the therapeutic mode, the patient's underlying condition or state is identified to the point at which an action can be taken. The clinical task is to determine what action should be taken.

Concern focuses on probabilities and values of possible outcomes, not on determining the underlying state of the problem. Treatment in this mode may confirm a diagnosis. In the diagnostic mode, emphasis is on determining the nature of the problem and identifying its causes. Treatment or action develops naturally from proper characterization of the problem (Elstein, 1977).

Diagnostic problem solving has been described as an iterative process consisting of four tasks: collection of information, generation of hypotheses, interpretation of evidence, and evaluation of hypotheses for a diagnostic decision (Elstein, Shulman, Sprafka et al., 1976).

Also related to the diagnostic and therapeutic modes are two other clinical activities: screening and follow-up. Screening is a process of focused data gathering directed at a specific problem. The amount of material collected is relatively small, and the client contact is brief. Follow-up involves observing the occurrence of anticipated outcomes, determining if new outcomes have resulted in changes in the nature of the system, and updating decisions of action based on the update (Elstein et al., 1977).

Elstein (1977) noted that clinical reasoning employs a "hypothetico-deductive" method of determining data collection. In this method preliminary hypotheses are generated, and the clinician seeks data to test them. By limiting the problem space, through the generation of a small number of working hypotheses (short-term memory store), the clinician simplifies the problem of the larger, long-term memory store. It was found that physicians consider from four or five up to six or seven hypotheses at one time. This number is "well below the number of hypotheses in the long-term memory store of any reasonably experienced physician and is evidence for the proposition that the size of the working memory is considerably smaller than the size of the long-term memory store" (Elstein, 1977, p. 38).

It has been emphasized that not all clinical tasks involve physicians and the medical profession. Likewise, it has been noted that "the clinical model is especially well suited to the study of reading clinicians because it is problem-initiated and problem-directed." However, it has been suggested that some untrained reading specialists may not really be able to use the clinical model in its entirety because "they lack the skills necessary to carry out the last two stages of the model" (interpretation of evidence, and hypotheses evaluation) (Elstein et al., 1977, p. 6).

Clinical Information Processes in Reading (CLIPIR)

The Clinical Information Processes in Reading (CLIPIR) study was conceived to explore the nature of clinical problem solving in reading diagnosis and remediation among reading clinicians and teachers and to test empirically the Inquiry Theory of Clinical Problem Solving as it applies to the field of reading.

Because of the exploratory nature of the research, it was necessary to design studies to establish piloting procedures, determine relevant variables, and construct a data base for the study of clinical behavior.

SIMCASE development.--To provide an empirical data base of the clinical problem-solving behavior of teachers and reading clinicians, simulated cases of reading disabilities were developed. A simulated case or set of data representing a child with a reading problem is referred to as a SIMCASE (Lee & Weinshank, 1976).

The SIMCASE development team followed these procedures:

1. Generation of list of problems which might be included in a case. Variables considered were (a) frequency of problems, (b) range of etiologies, and (c) alternative views of the reading process.
2. Selection of the problems for possible SIMCASE development from the list.
3. Review of case pool in M.S.U. Reading Clinic and M.S.U. reading diagnosis courses for purpose of identification of suitable simulation cases.
4. Development of Data Categories, all the data which might be requested on a case by clinicians, such as home and school background information, formal and informal test results, audio tapes of reading, etc.
5. Evaluation of overall case contents by senior clinicians.
6. Preparation of SIMCASEs, including audio tapes, writing samples, test results, etc.
7. Re-evaluation of the complete cases by the independent evaluation of two senior clinicians for internal consistency, fidelity and comprehensiveness of case contents.
8. Production of two versions of each SIMCASE: a manually based or "boxed" version and a computer-based version, consisting of all information that could be keyset into the IRT computer plus tapes and other materials to which the computer could refer clinicians.
9. Piloting of procedures and Data Categories (Cue List) using senior clinicians.
10. Evaluation of manually based and computer-based SIMCASEs by SIMCASE Selection and Evaluation Team and SIMCASE Development Team.

Observational Study, 1977

In 1977, a study was conducted on the interaction of eight "very senior" clinicians with eight SIMCASEs, four of which were alternate forms of the other SIMCASEs. The paid clinicians were selected on the basis of (a) recommendation of the local school administrators, (b) recommendation of the M.S.U. faculty, and (c) performance in SIMCASE verification.

The task for the clinicians studied was threefold: (1) given unlimited time, to select materials, using a cue list contained in the SIMCASE (subjects were encouraged to verbalize their thinking); (2) to write a diagnosis and suggested remediation based on material they had examined in the SIMCASE; and (3) in a debriefing session, to verbalize their rationale for cue selection and interpretation (stimulated recall) (Vinsonhaler, 1979b).

Outcomes, 1977

Results of the Observational Study, 1977 data indicated that by using an analysis of the diagnostic consistency of clinicians as they described the strengths and weaknesses of a SIMCASE, it was found that the senior reading clinicians studied lacked the consistency that might be considered necessary in medicine. Diagnostic consistency was measured by comparing clinicians' diagnostic reports for each SIMCASE with reports on alternate forms of the same SIMCASE and by comparing an individual clinician's diagnostic report with those of other clinicians diagnosing the same SIMCASE. The low agreement found in these comparisons suggested that clinical diagnosis in reading might be unreliable.

CLIPIR Application Exploratory Study in Educational Research (CAESER), 1977

During the summer of 1977, an exploratory study designed to apply the clinical problem solving theory in training teachers was conducted. Two primary questions were addressed: (1) "Could teachers be trained to diagnose in the same manner as an effective experienced reading clinician?" and (2) "Would increases in diagnostic performance be accompanied by increases in diagnostic memory?" (Sherman, 1979).

Results of a five-week study involving 36 students in a reading-diagnosis course tentatively indicated that (1) students did learn to diagnose in the same manner as the senior reading clinicians and that (2) increases in diagnostic performance were accompanied by increases in diagnostic memory (Gil, Hoffmeyer et al., 1979).

Implications, 1976-1977

The Observational Study, 1977 results seemed to indicate that some expert reading clinicians were less precise in both diagnosing and reporting diagnostic results than were clinicians in other fields such as medicine. However, the results of that study also suggested that the diagnoses and diagnostic reports of some expert reading clinicians do remain fairly consistent and yield high diagnostic commonality scores. As in medicine, these clinicians employ a hypotheses-directed approach to the clinical-inquiry process of problem solving. The application study (CAESER) seemed to indicate that with proper training involving SIMCASEs with feedback, clinicians, reading specialists, and classroom teachers can be taught to prepare diagnostic reports that are consistent and that yield high commonality scores.

Recent Studies

Recently completed and/or ongoing studies based on the theoretic structure of the Inquiry Theory have investigated the clinical problem-solving skills of experienced reading and learning-disability specialists and classroom teachers as they diagnose and then propose remediation for a variety of reading problems. The ultimate goal of these collaborative research efforts is to improve the instruction, evaluation, and performance of reading clinicians. Additionally, researchers are now attempting to explicate and refine the Inquiry Theory of Clinical Problem Solving, thereby increasing its predictive powers (Gil, Hoffmeyer et al., 1979; Stephens, 1978).

Summary

In the reading as well as in the psychological literature, conceptually unclear terms abound, thus confusing and complicating the study of how information is processed.

Research procedures in problem solving, however, seem to have a common relationship in realistic task environments (Shulman & Elstein, 1975). Such procedures include (1) total-task or process-tracing studies, which are concerned with the sequential character of gathering information in order to make a decision or judgment; (2) in-basket studies, in which the decision maker receives information "input" for making a decision; and (3) tab-item methods, which provide objective, reliable data through predesignation of choices of action, such as troubleshooting electronic equipment performance failure. The major concern of investigators using the information approach (total-task investigations, in-basket, and tab-item studies) is to observe the process of thinking and judgment in as actual a task environment as possible so as to conceptualize human behavior as task determined or related.

In a thorough review of a number of theoretical models and research methods on thinking, human judgment, and decision making, Shulman and Elstein (1975) determined that those studies of information processing, decision making, policy capturing, and lens model "have rarely been applied to the investigation of educational problems" (p. 32). They then proposed a variety of ways in which the models and methods they reviewed might be valuable in education. One major area of their discussion dealt with possibilities of research in diagnosing and remediating reading difficulties. One suggestion Shulman and Elstein made was to collect thinking-aloud protocols of experienced reading diagnosticians as they dealt with a series of cases. This research strategy, as well as others they suggested, is important because, "as in so many situations involving clinical judgment, the principles governing decision making are typically unclear." They concluded that it is crucial to examine <u>how</u> the problem solver "sizes up the situation, how the problem is formulated, what is judged to be relevant and what irrelevant, which sources of information are considered useful and which of no importance" (pp. 34-36).

While looking at the process of reading diagnosis, it was pointed out that there are problems inherent in diagnostic problem solving that should be considered. These interrelated problems include (1) the essence of diagnosis, the etiological and therepeutic aspects and the procedural steps; (2) the causal debate, singular or pluralistic and etiological or therapeutic; and (3) the lack of standardized terms. The basic disagreements over such issues have complicated communication both within and between professions (Whitecraft, 1971).

One effort to uncomplicate matters of clinical diagnosis in medicine and subsequently in reading has been the research relating to the Inquiry Theory of Clinical Problem Solving. A number of studies, each building upon the other, have been undertaken to examine clinical diagnosis in reading. These studies, designed to test the Inquiry Theory, hold promise for improving the instruction, performance, and evaluation of reading clinicians.

CHAPTER III

DESIGN AND PROCEDURES

Introduction

The primary objectives of this study were (1) to obtain data on the kinds of cues and the time and order in which cues were collected by each clinician in each clinical session, (2) to obtain objective data on the diagnostic statements made in a written diagnosis on each case by having clinicians transfer their written diagnostic statements to a standardized diagnostic check list, and (3) to obtain objective data on the use of hypothesis generation by having clinicians complete the Hypothesis/Observation Check List (H/OCL) and then transfer their responses to the standardized Reading Diagnostic Check List (RDCL). Secondary objectives were to obtain informal data from the clinicians in terms of (1) <u>what</u> information should be included in a good diagnosis and (2) <u>how</u> the clinicians themselves conduct a diagnosis.

Sample

The eight subjects for this study were chosen by university faculty recommendation from a list of reading clinicians who had taught the summer institute courses in reading diagnosis and remediation offered by Michigan State University. This procedure was used in an effort to select some of the most experienced clinicians in the mid-Michigan area. From those individuals recommended, eight volunteer clinicians (subjects) were selected. There were two male and six female clinicians. Six of the eight subjects either had, or were in the process of obtaining, a Ph.D. All were experienced classroom (5 to 7 years) and reading teacher/diagnosticians (3 to 15 years), and all had taught university courses in reading diagnosis and remediation. All of the clinicians were paid at a professional rate for their participation in the study. Biographical data on the clinician subjects are presented in Appendix A.

Research Design

The eight cases (four cases and their replicates) used in this study were based on children who had at one time been clients of the Michigan State University Reading Clinic. An attempt was made to select those cases that were described by a number of reading clinicians as being representative of the reading problems most frequently encountered in the public schools. The representative reading problems upon which the cases (SIMCASEs) were based included sight-word deficiencies, inadequate structural and phonetic analysis skills, inadequate fluency of oral reading, and poor comprehension. Data on four students were used in developing the eight SIMCASEs; each SIMCASE had two equivalent forms. Equivalent forms were prepared by making minor alterations in the original data base, such as changing the artist's sketch, the name, birthdate, father's occupation, or age and sex of siblings.

The available information for any particular SIMCASE included such data as family background, classroom information, achievement and intellectual-capacity tests, and individual and group reading measures. The information for the SIMCASEs was presented in five forms: as test scores, examiner's comments, test booklets, audio recordings, and test directions. Each of the cases contained initial contact information, which included an artist's drawing of the child based on his taped voice, an audio recording of an interview with the child, and brief background information.

The clinicians for the study participated in three, approximately three-hour sessions spaced no less than one week nor more than four weeks apart.

All stimulus materials, including SIMCASEs and equivalent forms, were subjected to counterbalancing to minimize systematic effects.

Subjects for the study were randomly assigned to the case order in the manner shown in Figure 5. A summary of the research design for this study is presented in Figure 6.

As shown in Figure 6, an attempt was made to balance the design in terms of easy and difficult cases. The decision about which cases were the easiest and which were more difficult was made by experienced reading clinicians who had worked with the students upon whom the cases were based. These clinicians seemed to agree that easy cases would be those that appeared most frequently. Difficult cases would be (1) those not seen often, (2) those in which the explanation of the problem was less obviously related to its cause(s), and/or

Condition	Subject #	Session 1	Case	Session 2	Case	Session 3	Case
1	107	d	1	D	3	d	1'
2	105	D	3	e	4	D	3'
3	108	E	2	d	1	E	2'
4	106	e	4	E	2	e	4'
5	103	d	1	D	3	d	1'
6	101	D	3	e	4	D	3'
7	102	E	2	d	1	E	2'
8	104	e	4	E	2	e	4'

E_2, $E_{2'}$ } Easiest Case and Replication

e_4, $e_{4'}$ } Second Easiest Case and Replication

d_1, $d_{1'}$ } More Difficult Case and Replication

D_3, $D_{3'}$ } Most Difficult Case and Replication

Figure 5.--Random assignment of subjects to clinical sessions.

(3) those in which there were more obscure answers to why the child couldn't read.

		Middle Case	
		D/d	E/e
First/Last Case	D/d	d D d'	D e D'
	E/e	E d E'	e E e'

Case Difficulty Scale

	E E'	e e'	d d'	D D'	
EASY					DIFFICULT
Case:	2	4	1	3	

Figure 6.--Research design.

Instrumentation and Data Collection

Since the analysis of the data involved the use of instruments created especially for this study, the problems of validity and reliability of the instruments must be considered. The problem of validity, i.e., whether the instrument measures what it purports to measure, was tested by means of a pilot study of the instruments and the instructions for their use. Instruments and instructions were found to be valid. Content validity was considered adequate justification for using the instruments developed for this study.

The question of reliability, i.e., how consistently the instrument measures whatever it measures, entails a check on the extent to which the instrument yields similar results upon repeated trials. Reliability of the instruments used in this study was a recognized concern of the researcher. However, since no standardized tests or instruments that purport to measure the decision-making and problem-solving behavior of reading clinicians were known to exist, instruments were developed for the study. Those instruments are the Reading Diagnostic Check List (RDCL) and the Hypothesis/Observation Check List (H/OCL) (See Appendix C).

Reliability

To measure the subject reliability on use of the RDCL, the clinicians were mailed an uncircled, carbon copy of each of their written diagnoses one week after the third clinical session. An accompanying letter (see Appendix E) instructed subjects to follow the same procedures used in the clinical sessions for circling and numbering their written diagnoses and then transferring them to the RDCL.

Table 1 shows the correlation (r_{xy}) for the clinic conversion and the home conversion of the written diagnoses, which had been done in the observational session, to the RDCL. Clinic conversion refers to the transfer by the subject in the clinical session of his written diagnosis to the RDCL. Home conversion refers to the transfer by the subject at home of his written diagnosis (done in the clinic and mailed to him) to the RDCL.

Table 1.--Correlation (r_{xy}) and Porter Index for clinic and home conversion of written diagnosis to RDCL on four cases.[a]

Subject No.	Case No.	Run No.	Pearson (r_{xy})	Porter Index
101	3	1	.57	.41
101	3'	3	.47	.32
101	4	2	.36	.24
102	2	1	.69	.54
102	2'	3	.68	.53
102	1	2	.65	.50
103	1	1	.50	.33
103	1'	3	.56	.41
103	3	2	.25	.15
104	4	1	.52	.36
104	4'	3	.56	.41
104	2	2	.55	.39
105	3	1	.15	.10
105	3'	3	.38	.28
105	4	2	.41	.41
106	4	1	.55	.36
106	4'	3	.53	.41
106	2	2	.40	.39
107	1	1	.44	.29
107	1'	3	.52	.38
107	3	2	.34	.24
108	2	1	.38	.25
108	2'	3	.62	.47
108	1	2	.34	.23

Note: Prime (') = alternate form for case.

[a]r to z transformation was <u>not</u> used because distribution was normal.

The Pearson product-moment coefficient was calculated on the relationship between the two measures of check list use. The Porter Index, a statistic developed by Andrew Porter, at Michigan State University, to analyze data in the National Day Care Home Study (see Wilcox, 1977), was also computed on clinic and home data in the present study. Results of the Porter Index analysis appear in the last column of Table 1. A more detailed discussion of the Porter statistic is presented in Chapter IV.

As can be seen in the table, the correlation ranged from .15 for Subject 105, Run 1, on Case 3, to .69 for Subject 102, Run 1, on Case 2. The reported values indicate that the clinicians in this study did not show a high degree of reliability in their use of the check list at Time 1 (clinic) and Time 2 (home).

Subject differences using mean and standard deviation for r_{xy} are shown in Table 2. The values indicate that subjects 102 and 105 were higher and lower, respectively, than the other clinicians in the group in terms of the reliability of their check-list use.

When case differences were considered, the variability was greatest for Case 3. This was influenced by the extreme value of .15 for Subject 105, Run 1, who diagnosed Case 3, the most difficult case.

Case differences using mean and standard deviation for r_{xy} are reported by case in Table 3. The values indicate that Case 3 had the least check-list reliability.

The variability in check-list reliability could have resulted from subject differences or case difficulty. Other possible reasons for the limited reliability of the check list across all cases and

Table 2.--Means and standard deviations for subjects based on correlation (r_{xy}) for use of RDCL.

Subject No.	$\underline{M}$ $\underline{r}$	SD
101	.47	.11
102	.67	.02
103	.44	.16
104	.54	.02
105	.31	.14
106	.50	.08
107	.43	.09
108	.45	.15

Table 3.--Means and standard deviations for cases based on correlation (r_{xy}) for use of RDCL.

Case No.	$\underline{M}$ $\underline{r}$	SD
1	.50	.11
2	.55	.14
3	.36	.15
4	.49	.08
	Mean total = .48	
	SD = .08	

subjects might have been the check list's complexity and/or its length (507 items), the lack of experimental control during the home conversion of written diagnoses to the check list, other factors, or combinations thereof.

Instruments

The RDCL was designed to objectify the data collection and analysis. Each clinician converted his written diagnosis to the RDCL, indicating strengths, weaknesses, or observations for diagnostic statements on the list corresponding to his written statements. This procedure was used for each session. The RDCL was developed from a Taxonomy of Reading Factors (TRF), which has been under development in the Clinical Studies Research Project of the Institute for Research on Teaching (IRT). In its present form, the RDCL includes 9 major categories and 169 subcategory items. In addition, there is a tenth category called "Other," and within each of the nine categories is an "Other" subcategory. The statements for "Other" were not included in the data analysis; they are reported separately in Appendix D.

The size of the RDCL (169 items with 3 possible responses per item, for a total of 507 possible responses) was another concern in the study. Earlier attempts had been made (Vinsonhaler, 1979b) to reduce the size of a similar check list by combining categories, since most agreement measures are sensitive to vocabulary size. However, combining the categories yielded "large numbers of inconsistencies (e.g., the same diagnosis often includes inconsistent statements such as: 'no problem with phonics' and 'needs work on long vowels')" (p. 14).

The H/OCL (Hoffmeyer, 1979) was designed to objectify and simplify the data collection and analysis. Using the H/OCL, the clinician subject was asked to respond to two questions regarding his request for information (cues) on a case (SIMCASE) by selecting his answers from a list of responses to the questions "Why did you ask for this piece of information?" and "What did it tell you?" The clinician was also asked to write an explanation for each response he checked. The researcher later applied these data to the RDCL, to compare subjects' responses more objectively. The subject was not asked to make the conversion because of the additional session time that task would have required. In converting the H/OCL (debriefing data) to the RDCL, every effort was made to represent the clinician's intentions by referring to his written diagnosis for clarification.

Statement Concerning Hypotheses

Because of the exploratory nature of this study, research hypotheses were not stated. Rather, this research was concerned with formulating a systematic description of the reasoning of reading clinicians as they diagnosed simulated reading-problem cases.

The questions addressed were investigated through both formal and informal measures designed to provide information on <u>how</u> experienced reading clinicians diagnosed specific reading problem cases, <u>what</u> information they used in making a diagnosis, and <u>why</u> they asked for certain information. The formal questions concerned the agreement of eight clinicians on the data (cues) they collected and the diagnostic statements they made when diagnosing reading problem cases and whether the clinicians used hypotheses to direct their clinical inquiry in arriving at a diagnosis. The informal questions pertained to the clinicians' own perceptions of <u>what</u> information should be included in a good diagnosis and <u>how</u> the clinicians went about making a diagnosis.

Procedures

The procedures for the three sessions in this study were as follows:

1. A background questionnaire was completed by clinicians and followed by instructions with practice on a sample SIMCASE.
2. Initial contact information was provided on the SIMCASE.
3. Data were collected by the clinicians <u>without</u> an inventory of data available; directions were given and a recording was made by the examiner.
4. Each clinician wrote a diagnosis and a remediation report.
5. The clinicians then transferred their written diagnoses to the RDCL.
6. Clinicians recalled their reasons for data collection by responding to a written questionnaire (stimulated recall).
7. Oral questions regarding the last case were asked in the third session <u>only</u>.

Taking one subject at a time for each session, each individual in the study was asked to complete a background questionnaire. (See Appendix A.) Then the examiner gave the clinician detailed instructions and had him practice using a sample SIMCASE and sample data identical in form to the ones he would use in the sessions. Next, the clinician was given some initial contact data on the case: (1) an artist's sketch of the student based on taped voice, (2) a taped interview with the student, and (3) written background information on the student.

The subject was asked to request information on the SIMCASE and to use the data he requested in making his diagnosis. All requests were recorded on audio tape, and the examiner noted on the H/OCL the time of the request and the name of the requested information.

The SIMCASE was a simulated case or set of information representing a child with a reading problem. SIMCASE materials were contained in a small file box, and the examiner provided the information by handing the clinician the requested information. The clinician did not know what materials were in the SIMCASE; i.e., he was not given a cue list. Subjects were given 45 minutes to request and use SIMCASE materials. Subjects could take notes if they wished, but they were not required to do so.

Following the SIMCASE/clinician interaction, the clinician was asked to write, in sentence form, a diagnosis and remediation for the SIMCASE in the way he would usually write a diagnosis and remediation. Special carbon paper was used, to make a duplicate copy of the subject's handwritten diagnosis and remediation. The clinician was given 25 minutes to write a diagnosis and 25 minutes to write a remediation.

Next, the clinician was shown a written diagnosis of a sample case for which diagnostic statements had been circled and numbered. The subject then practiced the task with another sample diagnosis. The clinician was subsequently given a sample RDCL and was instructed in its use. This was followed by practice with the sample check list in transferring the circled and numbered sample diagnosis to the RDCL sample.

The clinician was then asked to circle and number, on one copy of his written diagnosis, all diagnostic reading statements. He then transferred these statements to the RDCL. The number from a written diagnostic statement was placed in the appropriate place on the check list, as either a strength, a weakness, or an observation, beside the statement on the check list that corresponded most closely to the clinician's own statement. An observation was defined as a nonvalue or neutral statement; e.g., "The student has brown hair and brown eyes." The circling and numbering of statements and the transfer to the check list were <u>not</u> timed tasks.

During a short break for the subject, the examiner arranged the materials the subject had requested, in the order in which they had been requested. After the break, the subject was asked to complete the H/OCL, responding to questions of <u>why</u> he had asked for a particular piece of information (cue) and <u>what</u> the information had told him (hypotheses, observations, or hunches). The examiner recorded the order in which cues had been requested; the actual materials requested provided the stimulus for recall. This was <u>not</u> a timed task.

The above procedures were followed for each of the three clinical encounters, except that in the third or last session the clinicians were asked to respond to a short oral questionnaire regarding the third case. (See Appendix C.) The questions dealt with the clinician's opinion about the student's ability in specific areas of reading. Also, the questionnaire was designed to determine if the clinicians used a "model" in the diagnostic process (Sherman et al., 1978).

Summary

The major objective of this study was to determine the nature and extent of agreement of eight experienced reading clinicians as they collected cues, made diagnostic statements, and formulated hypotheses relative to making a diagnosis on simulated reading-problem cases.

The eight clinician subjects were asked to interact with simulated materials representing a reading problem case and then to write a diagnosis and a remediation. The written diagnosis was transferred to the RDCL for objectification of the data. Responses to questions of <u>why</u> certain case data were requested on each case and <u>what</u> information was provided by those data were also applied to the RDCL.

Additional procedures for the third or last session included questions designed to (1) elicit information regarding the clinician's conception of the content or data base of a "good" diagnosis, (2) determine the clinician's diagnostic routine, and (3) determine the source or schema that allowed the data base (memory) to be translated into action (strategy). Other questions dealt with four areas of reading and how they could be defined for the third or last case each subject diagnosed. The four areas were (1) instant word recognition, (2) word analysis, (3) reading fluency, and (4) reading comprehension.

All stimulus materials used in the study, including cases and equivalent forms, were subjected to counterbalancing to minimize systematic effects. Subjects were randomly assigned to the case order.

Subject reliability on the use of the RDCL, a previously unused instrument, was measured using a test-retest procedure. The Pearson product-moment correlation coefficient ($\underline{r}$) was calculated as the measure of relationship. The value of $\underline{r}$ ranged from .15 to .69. The Porter statistic, $\frac{A}{A+B+C}$, an index of the proportion of agreement that excludes clinician agreement not to select a cue or make a diagnostic statement, was also calculated on the same data. The Porter values ranged from .10 to .54.

CHAPTER IV

ANALYSIS MEASURES AND RESULTS

The measures of clinical problem-solving behavior used in this research were developed for this study as well as for the larger research project of which it is a part. Earlier work on problem solving in medical diagnosis provided the framework for determining the appropriateness of the measures for investigating clinical problem-solving behavior in reading diagnosis (Norman, 1977; Elstein et al., 1976).

Because of the limited sample size used in this study and the fact that the methodology and statistical measures are still open to speculation and investigation, all findings and conclusions offered here must be considered tentative.

Two major types of formal measures were used in the study reported here: (1) the <u>product</u> measures, which were intended to measure the outcomes of the clinical interaction between a reading clinician and a case (SIMCASE); and (2) the <u>process</u> measures, which were intended to measure the way in which the reading problem was diagnosed.

A third type of measure used in the study was informal and dealt with clinic interviews, which pertained to clinicians' diagnostic opinions regarding certain aspects of reading diagnosis. No formal or statistical measures were employed to analyze the interview data. Rather, simple descriptive techniques were used.

Product Measures

Certain key behaviors govern the clinical interaction between a clinician and a case. The product measures were used to account for the _results_ of the interaction, i.e., the data collected and the written diagnostic statements made by a given clinician to diagnose a specific case. The analysis of the resulting data is divided into four parts:

1. Proportional agreement, which is a measure of group agreement on cues collected and diagnostic statements made on the same case or an alternate form of the case;

2. Commonality, which is a measure of agreement between an individual clinician and a defined group of clinicians in terms of cues collected and diagnostic statements made on the same case or its alternate form;

3. Inter/intraclinician agreement, which is a measure of the agreement between one clinician and another clinician (or one clinician with himself) on the cues collected or the diagnostic statements made on the same case or on an alternate form of the case; and

4. Intraclinician agreement, which is a measure of the agreement of one clinician's cue collection for a case with his own cue collection for an alternate form of the case. The same analysis applies to diagnostic statements made.

The Proportional-Agreement Statistic

The proportional-agreement statictic provides data on the similarities between the cues collected (or diagnostic statements made) by a group of clinicians for a given case. To determine similarities among cues and among statements, a standard for comparisons was first established. The standard developed for cues was the cue domain of data available on each particular SIMCASE. (See Appendix F.) The standard developed for statements was the statement domain or RDCL for all four cases. The statement domain was composed of categories and subcategories of diagnostic reading factors designed to encompass the spectrum of reading problems. The computation formula appears in Appendix I.

Results of proportional agreement for cues collected.--Table 4 contains the results for the proportional agreement statistic for the cues most frequently requested (by at least three of the six clinicians) for Case 1. Data for the other three cases appear in Tables H1, H2, and H3 of Appendix H. A sample cue domain for one case appears in Appendix F.

Proportional agreement for a given cue was the proportion of those clinicians who diagnosed a particular case and who requested that same cue. For example, for Case 1 (Table 4), cue number 17, "Durrell Silent Reading (Test Booklet)," was requested by roughly 83 percent or five of the six clinicians diagnosing that case. It should be recalled that the clinicians did not know what cues or data sources were available for a case.

Table 4.--Proportion of clinician agreement on most frequently collected cues, Case 1.

Cue Number	Proportion Value	Cue Name	
4	.83	BKG 20	Classroom Information (EC)
7	.67	DOL 3	Dolch Vocabulary Word List (TB)
12	.83	DUR 3	Durrell Oral Reading (TB)
13	.67	DUR 4	Durrell Oral Reading (AR)
17	.83	DUR 9	Durrell Silent Reading (TB)
21	.67	DUR 15	Durrell Listening (TB)
25	.50	DUR 21	Durrell Word Recognition and Word Analysis (TB)
30	.50	DUR 27	Durrell Visual Memory of Words--Primary (TB)
43	.50	GMG 1	Gates-MacGinitie Vocabulary (TS)
53	.67	GMK 3	Gates-McKillup Blending Common Word Parts (TB)
58	.67	INF 4	Informal Oral Reading (AR)
68	.50	PEA 9	Peabody Reading Comprehension (TB)
81	.50	WISC 7	Weschler Verbal (TS)
84	.50	WISC 13	Weschler Performance (TS)

Key: TS = Test Scores; EC = Examiner's Comments; TB = Test Booklet; AR = Audio Recording; TD = Test Directions; Total available cues = 87

As indicated in Table 4, the most frequently requested cues for Case 1 were Examiner's Comments on Classroom Background Information (.83) and the Durrell Oral Reading (.83) and Durrell Silent Reading (.83) Test Booklets. Fourteen cues out of a possible 87 total cues available for Case 1 were requested by three or more clinicians. The most frequently requested form of information was Test Booklet (TB), which provided the actual test items. The clinicians diagnosing Case 1 were apparently interested in seeing the kinds of errors the student made.

The percentage of agreement on the most frequently collected cues (by three or more clinicians) is indicated, by case, in Table 5. The reported data tend to convey more agreement than actually occurred because only the <u>most frequently</u> requested (.50 to 1.0) cues were used to compute the proportion of agreement. The percentages in <u>total</u> proportion of agreement (P.A.) for cues on four cases appear in Table H4 of Appendix H. As shown in that table, between 42 and 52 percent (an average of 48 percent) of the cues available were not collected on each of the four cases. An average of only 23 percent of the total cues available for Cases 1 through 4, respectively, was agreed upon by three or more (.50-1.0) of the six clinicians diagnosing each case.

<u>Results of proportional agreement for diagnostic statements</u>.--Table 6 shows results of proportional agreement for the diagnostic statements most frequently used in all four cases when the written diagnoses were converted to the standardized check list. (See Appendix C for the diagnostic statement domain check list for the

Table 5.--Percentages for proportion of agreement on cues most frequently requested for four cases.

Case Number	Number of Cues Collected	3 Sessions (P.A.=.50)	4 Sessions (P.A.=.67)	5 Sessions (P.A.=.83)	6 Sessions (P.A.=1.0)
Case 1	14	43%	36%	21%	0%
Case 2	24	29%	38%	33%	0%
Case 3	27	44%	22%	26%	7%
Case 4	21	48%	14%	14%	19%

four cases.) The check list contained 169 diagnostic statements with three evaluative choices--strength, weakness, or observation--for each statement. Therefore, the check list domain had a total of 507 items. Proportional agreement for a given diagnostic statement was the proportion of those clinicians who agreed by mentioning that statement. For example, "Potential for grade-level work--reading (strength)" was mentioned by 50 percent (three out of six) of the clinicians for Case 4, was not mentioned for Case 2, and was mentioned only once each for Cases 1 and 3.

The first column in Table 6 lists the diagnostic statements from the check list that were used in 50 percent (three out of six) of the diagnoses for a single case. The last four columns indicate, by case, the proportion of clinicians making each statement. For example, the first diagnostic statement, "Word recognition--general (weakness)," was used by 50 percent of the clinicians for Case 1, 33 percent for Case 2, 33 percent for Case 3, and 50 percent for Case 4.

Table 6.--Diagnostic statements most frequently selected from a standard check list.[a]

Diagnostic Statement From Standard Vocabulary Checklist		Proportion of Diagnoses Including Statement: Arrayed by Case			
		Case 1	Case 2	Case 3	Case 4
Word recognition--general	W	.50	.33	.33	.50
Rate of reading--silent	W	.50	.33	.33	.17
Intell/ed. potent.--general	S	.33	...	.50	.83
Progress in school--reading	W	.50	.33	.50	.50
Comprehension--general	W	...	.50	.67	.17
Hearing--acuity	W	...	.50	.33	...
Attitude toward reading--independent	W	.50	...	.17	.33
Basic sight words--score	W	.33	...	...	.67
Word analysis--general	W	...	.33	.17	.67
Phonetic analysis--general	W	.17	.17	.33	.83
Use of suffixes	W	...	.50	...	.33
Rate of reading--oral	W	.50	.67	...	...
Word recognition--basic sight word	W	.33	...	...	.83
Vision--general statement	S	...	...	...	.50
Verbal intellectual poten.	S	.17	...	.17	.50
Potential for grade-level work--reading	S	.17	...	.17	.50
Emotional adjust.--general	W	...	...	.50	...
Visual discrim.--whole word	W	.50	...	...	.17
Word recognition--general	O	.17	.50	.17	...
Word recognition--basic sight word	S	...	.50	.17	...
Use initial consonant sounds	W	...	.67	...	.17
Use of blends--specific	W	.17	.50	...	.17
Comprehension--general	S	.83	.17	...	.17
Comprehension--general	O	...	.50	...	...
Oral reading--general	W	...	...	.50	...

[a]Statements mentioned in 50 percent of the diagnoses for a single simulated case, or in the diagnoses for 50 percent of the cases, or both. Total diagnoses per case = 6.

Key:
S = strength ($\underline{n} = 6$)
O = observation ($\underline{n} = 2$)
W = weakness ($\underline{n} = 17$)

Proportion		Diagnoses
.83	=	5
.67	=	4
.50	=	3
.33	=	2
.17	=	1

The data presented in Table 6 seem to indicate that the most frequently selected diagnostic statements over all cases were, equally, "Word recognition--general (weakness)" and "Progress in school--reading (weakness)." However, the statements with the highest agreement on any single case were "Intellectual/educational potential--general (strength)" for Case 4, "Phonetic analysis--general (weakness)" for Case 4, "Word recognition--basic sight words (weakness)" for Case 4, and "Comprehension--general (strength)" for Case 1, each having 83 percent agreement or being selected in five out of six sessions for the particular case.

The percentages in proportion of agreement for the most frequently mentioned diagnostic statements made in three (.50), four (.67), five (.83), and six (1.00) sessions are indicated, by case, in Table 7.

Table 7.--Percentages in proportion of agreement for most frequently mentioned diagnostic statements[a] on four cases.

Case Number	3 Sessions (P.A.=.50)	4 Sessions (P.A.=.67)	5 Sessions (P.A.=.83)	6 Sessions (P.A.=1.0)
Case 1	24%	0%	4%	0%
Case 2	28%	8%	0%	0%
Case 3	16%	4%	0%	0%
Case 4	20%	8%	12%	0%

Most frequent statements for four cases = 25.

[a]No statements were mentioned in all six sessions for any case.

The results of proportional agreement for statements indicated that clinicians in this study had only slight agreement as a group on the diagnostic statements made for a given case. The low agreement for diagnostic statements might have been a result of the low reliability of the check list (.48), the length of the check list, and/or the number of check-list categories used to determine agreement.

The percentages in <u>total</u> proportion of agreement (P.A.) on diagnostic statements for four cases appear in Table H5 of Appendix H. As shown in the table, between 77 and 82 percent (an average of 80 percent) of the diagnostic statements appearing in the check list were not checked for the four cases. An average of only 1 percent of the same diagnostic statements in the check list were checked by three or more (.50-1.0) of the six clinicians diagnosing each case.

The Commonality-Agreement Score

The commonality score is intended to reflect the agreement of one clinician's collection of cues with the cues collected by a given group of clinicians on the same case (Vinsonhaler, 1979b). This score indicates the comparison between an individual clinician's cue collection and a group of cues collected by all other clinicians for the same case. It accounts for which cues are collected by each clinician, and how many are collected. The same score can be calculated for diagnostic statements.

Results of commonality agreement.--The values for commonality agreement on cues collected and diagnostic statements made in six diagnoses on each of four simulated reading cases are presented in Table 8. The data indicate that there was approximately 75 percent commonality or individual-to-group agreement on cues collected across all subjects and all cases. A comparison of the means for the individual cases indicated little variability between cases. The commonality scores seemed not to be affected by case difficulty as both an easy case (Case 4) and a difficult case (Case 3) had the same average commonality score. The data in Table 8 also indicate that there was approximately 55 percent commonality agreement on diagnostic statements made across all clinicians and all cases.

Table 8.--Commonality agreement on cues collected and diagnostic statements made in six diagnoses on each of four cases.[a]

Case Number	Cues			Diagnostic Statements		
	M	Range	*SD*	*M*	Range	*SD*
1 (*n*=6)	.71	.48/.82	.13	.55	.37/.76	.16
2 (*n*=6)	.74	.30/.93	.24	.54	.25/.66	.15
3 (*n*=6)	.77	.39/.95	.20	.47	.28/.64	.15
4 (*n*=6)	.77	.62/.86	.11	.62	.42/.78	.12
Mean total =	.75		Mean total =	.55		
SD =	.03		*SD* =	.06		

[a]Commonality is bounded by 0 and 1.

Results of calculations shown in the commonality table suggested that when individual clinicians were compared with the group of clinicians diagnosing the same case, there was greater agreement, on the average, among clinicians in this study on cues collected than on diagnostic statements made. Comparing cues and statements by case, it can likewise be shown that for any given case there was more commonality agreement on cues collected than on diagnostic statements made. Two tables of commonality (cues and diagnostic statements) showing frequency for intervals appear in Appendix H (Tables H6 and H7). Table 8 is a summary of the two commonality tables in the appendix.

Inter/Intraclinician Agreement

There are several ways to describe inter/intraclinician agreement for the cues collected and the diagnostic statements made in the clinical encounter. In this study, two different indices (the Phi coefficient and the Porter Index) were used, each providing somewhat different information. In each instance, the value reported is by case and for the clinicians diagnosing a particular case or its alternate form. The Phi, denoted by ϕ, is the traditional coefficient of correlation for nominal dichotomous data. One Phi coefficient was computed for each pair of clinicians. (See Appendix I.)

When interpreting the results, it should be noted that because of an oversight in doing the statistical analysis for interclinician agreement, the pairs of scores for intraclinician agreement were also included. The interclinician analysis for each case should only have made calculations for subjects compared with other subjects diagnosing the same case or its alternate form, <u>not</u> for the subject compared with himself. The inclusion of the <u>intra</u> data (two comparisons per case) along with inter data in the analysis of agreement might possibly have raised interclinician agreement slightly since intraclinician agreement was usually somewhat higher than interclinician agreement. The data were not reanalyzed because the Observational Study of 1977, used for comparison of the data in this study (see Chapter V), had the same analysis. The data for inter/intraclinician agreement in both studies will be reanalyzed later to exclude the intra portion. Comparisons of an individual to other individuals will be referred to as inter/intraclinician agreement in the dissertation text.

The Porter Index, the second of the two indices used in this study to describe inter/intraclinician (and intraclinician) agreement, was $\frac{A}{A+B+C}$, which describes the proportion of agreement when the base was the total number of cues collected (or diagnostic statements made) for which one or the other or both clinicians collected the cue or made the diagnostic statement (see Wilcox, 1977, pp. 54-60). The upper bound of the index would be the value of the index $\frac{A+D}{A+B+C+D}$, which describes the proportion of agreement when the base was the total number of cues in the particular cue domain (or diagnostic statements in the statement domain). The $\frac{A}{A+B+C}$ index excludes clinician agreement not to request a cue or make a diagnostic statement, i.e., the "d" cell (--) in the 2 x 2 contingency table. In general, the values of the Porter Index would be expected to be somewhat lower than the values of Phi when the "d" cell (--) is large.

Results of inter/intraclinician agreement.--The data presented in Table 9 show results of calculations for inter/intraclinician agreement on cues collected for four simulated reading cases. Both of the indices reported are for the same data. Table 9 is a summary of

Table H8, which appears in Appendix H. Table 9 indicates that there was consistently low agreement on cues collected for all four cases. Results of calculations using the two statistics do not appear to differ greatly, as indicated by the means and standard deviations for individual cases and totals. In considering case difficulty, it should be noted that the easiest case (Case 2) and the most difficult case (Case 3) had mean Phis of .35 and .38, respectively.

Table 10 is a summary of Table H9 in Appendix H. The summary table shows the means and standard deviations for inter/intraclinician agreement on diagnostic statements for four cases. The statistics used were the Phi coefficient and the Porter Index. As indicated in the table, there was consistently low inter/intraclinician agreement on diagnostic statements for the four cases, with slight differences between the values for the two measures reported. The data indicate very low to almost no agreement (Case 3) for subjects on diagnostic statements made for the four cases. Case 3 was the most difficult of the four cases and may account for the lower diagnostic agreement for that case.

A comparison of data for cues and diagnostic statements across all clinicians and all cases indicates lower agreement for diagnostic statements made than for cues collected.

Table 11 shows the intraclinician agreement statistics (Phi and Porter), by subject and case, for cues collected on four cases. The third column on the left side of the table shows the value for the Phi coefficient for each subject, when compared with himself, in terms of cues collected, on alternate forms of the same SIMCASE. The first

Table 9.--Means and standard deviations for inter/intraclinician agreement (Phi and Porter) on cues for four cases.

Case Number	Range	Phi Coefficient		Range	Porter Index	
		M	SD		M	SD
1 (n=15)	-.04/.55	.30	.18	.09/.46	.28	.12
2 (n=15)	-.18/.87	.35	.28	.05/.83	.36	.20
3 (n=15)	.04/.63	.38	.15	.09/.54	.33	.12
4 (n=15)	.21/.60	.42	.11	.23/.52	.37	.08
	Mean total =	.36		Mean total =	.34	
	SD =	.05		SD =	.04	

Table 10.--Means and standard deviations for inter/intraclinician agreement (Phi and Porter) on diagnostic statements for four cases.

Case Number	Range	Phi Coefficient		Range	Porter Index	
		M	SD		M	SD
1 (n=15)	.00/.28	.10	.08	.02/.17	.07	.05
2 (n=15)	-.03/.43	.12	.11	.00/.29	.09	.07
3 (n=15)	-.05/.26	.07	.08	.00/.16	.06	.04
4 (n=15)	.04/.31	.16	.07	.05/.20	.11	.05
	Mean total =	.11		Mean total =	.08	
	SD =	.04		SD =	.02	

column on the right side of the table provides the values for the Porter statistic on the same data used to compute the Phi coefficients presented in the third column of the table.

Table 11.--Means and standard deviations for intraclinician agreement for cues collected on four cases.

Case Number	Subject Number	Phi Coefficient	*M*	*SD*	Porter Index	*M*	*SD*
1	103 107	.05 .42	.24	.26	.13 .36	.25	.16
2	102 108	.50 .87	.69	.26	.42 .83	.63	.29
3	101 105	.43 .38	.41	.04	.32 .38	.35	.04
4	104 106	.60 .39	.50	.15	.52 .36	.44	.11
	Mean total =	.46			Mean total = .42		
	SD =	.19			*SD* = .16		

1 = More Difficult 3 = Most Difficult
2 = Easiest 4 = Second Easiest

As the data in Table 11 suggest, there was considerable variability among clinicians in this study in self- or intraclinician agreement on cues collected. The Phi ranged from .05 for Subject 103 on Case 1 (difficult) to .87 for Subject 108 on Case 2 (easiest). These differences might be case specific (i.e., relating to case difficulty) or a result of subject differences, since the other reported

Phi for Case 1 was .42, which was more in line with the values for the other cases.

Additionally, the data for intraclinician agreement on cues indicate that, on the average, the reading clinicians in this study showed only limited agreement with themselves on the cues they collected on alternate forms of the same case. The notable exception was Subject 108 for Case 2, who had a Phi value of .87 in agreement with himself on cues collected for the same case, alternate forms. This would indicate that Clinician 108 collected 87 percent of the same cues at Time 1 and Time 2 for the same case (alternate forms). However, since Case 2 had been determined to be the easiest of the four cases, one can speculate about the effect case difficulty might have had on Clinician 108's self-agreement.

Although it has not been determined what the value of Phi _should_ be clinically, statistically the low average intraclinician agreement on cues collected (Phi = .46) seems to indicate that, in general, the clinicians in this study followed no particular diagnostic routine or strategy in collecting information or data on a case in order to arrive at a diagnosis. However, individual clinicians, 108 for example, did appear to have a strategy or routine, as reflected in the agreement Phi for cues (Phi = .87). Case difficulty, however, might be an intervening variable that affected the value of Phi in this study.

Except for Case 1, the values of the Porter Index appear to be somewhat lower than the values reported for Phi.

Table 12 shows the intraclinician agreement (Phi and Porter) on diagnostic statements for both the clinic and the home conversion of diagnoses to the RDCL for four cases. The reliability of the two conversions (clinic and home) of the written diagnosis to the RDCL was reported in Chapter III, the mean reliability being .48 for the two conversions (.49 if z scores were used).

Table 12.--Intraclinician agreement for two conversions (clinic and home) of written diagnoses to the RDCL.

Case Number	Subject Number	Phi Coefficient				Porter Index			
		Clinic	Home	M	SD	Clinic	Home	M	SD
1	103	.28	.40	.34	.08	.17	.26	.22	.06
	107	.26	.31	.29	.04	.17	.21	.19	.03
2	102	.43	.41	.42	.01	.29	.27	.28	.01
	108	.21	.22	.22	.01	.14	.14	.14	.00
3	101	.26	.30	.28	.03	.16	.20	.18	.03
	105	.01	.12	.07	.08	.03	.09	.06	.04
4	104	.31	.36	.34	.04	.20	.23	.22	.02
	106	.26	.46	.36	.14	.17	.32	.25	.11

Table 13 shows the intraclinician diagnostic statement agreement statistics for the diagnosis of four simulated reading cases. In general, the data indicate that the reading clinicians for this study showed only limited agreement with themselves on diagnostic statements made for the same case. Additionally, it can be stated that although the overall mean intraclinician agreement Phi (.26) and Porter (.17) were limited, they were higher than the overall group mean or inter/intraclinician agreement Phi (.11) and Porter (.08) on statements. (See Table 10.) Thus, it appears that the reading clinicians in this study showed more agreement with themselves on diagnostic statements made for the same case than the group as a whole agreed with each other; i.e., the individual intra agreement was higher than the group inter/intra agreement.

Table 13.--Means and standard deviations for intraclinician agreement for diagnostic statements on four cases.

Case Number	Subject Number	Phi Coefficient			Porter Index		
			M	*SD*		*M*	*SD*
1	103 107	.28 .26	.27	.01	.17 .17	.17	.00
2	102 108	.43 .21	.32	.16	.29 .14	.22	.11
3	101 105	.26 .01	.14	.18	.16 .03	.10	.09
4	104 106	.31 .26	.29	.04	.20 .17	.19	.02
		Mean total = .26			Mean total = .17		
		SD = .08			*SD* = .05		

In addition, when the values of intra Phi for statements (Table 13) are compared with the values of intra Phi for cues (Table 11), it can be seen that, on the average, subjects for this study agreed more with themselves on the cues they collected (.46) for the same case (alternate forms) than they agreed with themselves on the diagnostic statements made for the same case (alternate forms).

Looking at the data in Table 13 for each subject, it appears that Subject 105 had almost no agreement with himself on the diagnostic statements made for the same case at Time 1 and Time 2. Again, as with the intra agreement data on cues (Table 11), the low agreement on statements may be attributable to subject (clinician) differences or to case difficulty, since Subject 105 diagnosed the most difficult case. It should be noted that the lowest intra agreement Phi value (.05) for cues was for Subject 103, who diagnosed the first difficult case (Case 1), whereas the lowest intra agreement Phi value (.01) for statements was for Subject 105, who diagnosed the most difficult case (Case 3). The highest Phi value (.43) for intra agreement on statements was for Subject 102, who diagnosed the easiest case (Case 2). The highest Phi value (.87) for intra agreement on cues collected was for Subject 108, who also diagnosed the easiest case (Case 2).

The Inquiry Theory states that the greater the similarity of clinical memory (problem, cue values, prescription and treatment descriptions, and the relations between them), the greater the agreement of diagnoses. One descriptive part of this theory represented by the agreement corollary states that the agreement in cues collected by the same individual diagnosing alternate forms of the same case is greater than or equal to comparisons made between individuals (the group) diagnosing alternate forms of the same case. (See Chapter I.) Therefore, if the corollary holds, the intraclinician-agreement measures should yield equal or higher values than the inter/intra-clinician-agreement measures.

A summary of the agreement statistics on cues collected for four cases is presented in Table 14. The data indicated that when the inter/intraclinician Phi was compared by case to the intraclinician Phi, the intraclinician values were higher for Cases 2, 3, and 4. However, Case 1 showed a higher mean Phi for inter/intraclinician agreement (.30) than for intraclinician agreement (.24). Therefore, it would appear that for cue agreement, the corollary held for Cases 2, 3, and 4 but did <u>not</u> hold for Case 1. In terms of the Inquiry Theory, it would seem that the clinicians diagnosing Case 1 did not share a memory for cues. Whether or not this variability is a result of case difficulty and/or subject differences is not clear. One might question what is unique about Case 1 and/or the subjects diagnosing that case.

The second corollary of the Inquiry Theory postulates that the common elements between diagnostic statements made for two forms of the same case by the same clinician should be greater than or equal to diagnoses prepared by different clinicians on the same case. This would indicate that the intra agreement for diagnostic statements should be equal to or higher than the inter/intraclinician agreement.

A summary of the agreement statistics for diagnostic statements on four cases is presented in Table 15. As the data indicate, the intraclinician (individual) agreement for diagnostic statements was higher than the inter/intraclinician (group) agreement for diagnostic statements on all four cases. Therefore, the corollary held and the clinicians shared a memory, albeit a limited one, for diagnostic statements.

Table 14.--Agreement statistics on cues collected for four simulated reading cases.

Statistic for Agreement	Cues Collected				Average Over Four Cases
	Case 1	Case 2	Case 3	Case 4	
No. clinicians	4	4	4	4	
No. diagnoses	6	6	6	6	
Commonality score					
Mean	.71	.74	.77	.77	.75
Std. dev.	.12	.24	.20	.11	.03
Inter/intra diagnosis correlation (Phi)					
Mean	.30	.35	.38	.42	.36
Std. dev.	.18	.28	.15	.11	.05
Inter/intra diagnosis (Porter)					
Mean	.28	.36	.33	.37	.34
Std. dev.	.12	.20	.12	.08	.04
Intra-diagnosis correlation (Phi)					
Mean	.24	.69	.41	.50	.46
Std. dev.	.26	.26	.04	.15	.19
Intra-diagnosis (Porter)					
Mean	.25	.63	.35	.44	.42
Std. dev.	.16	.29	.04	.11	.16

Table 15.--Agreement statistics for diagnostic statements made on four simulated reading cases.

Statistic for Agreement	Diagnostic Statements Made				Average Over Four Cases
	Case 1	Case 2	Case 3	Case 4	
No. clinicians	4	4	4	4	
No. subjects	6	6	6	6	
Commonality score					
Mean	.55	.54	.48	.62	.55
Std. dev.	.16	.15	.15	.12	.06
Inter/intra diagnosis correlation (Phi)					
Mean	.10	.12	.07	.16	.11
Std. dev.	.08	.11	.08	.07	.04
Inter/intra diagnosis (Porter)					
Mean	.07	.09	.06	.11	.08
Std. dev.	.05	.07	.04	.05	.02
Intra-diagnosis correlation (Phi)					
Mean	.27	.32	.14	.29	.26
Std. dev.	.01	.16	.18	.04	.08
Intra-diagnosis (Porter)					
Mean	.17	.22	.10	.19	.17
Std. dev.	.00	.11	.09	.02	.05

Product Measures: Summary

The product measures data presented in this chapter seem to suggest the following:

1. Twenty-three percent of the same cues from the total cues available on the four cases were collected by three or more (P.A. = .50-1.0) of the six clinicians diagnosing each case.

2. Fifty-two percent of the cues available were collected across all four cases.

3. One percent of the <u>same</u> diagnostic statements in the check list were checked by three or more (P.A. = .50-1.0) of the six clinicians diagnosing each case.

4. Twenty percent of the diagnostic statements in the check list were checked for the four cases.

5. The commonality scores indicated that the clinicians for this study were consistently high, 75 percent, in their agreement on cues collected across all cases and all subjects.

6. The commonality scores also indicated that the subjects for this study agreed, on the average, 55 percent of the time on the diagnostic statements made across all cases and all subjects.

7. The inter/intraclinician agreement Phi for cues indicated little variability between cases on cues collected (M = .36, SD = .05).

8. The inter/intraclinician agreement Phi for diagnostic statements indicated little variability between cases (M = .11, SD = .04) but very low agreement.

9. The intraclinician agreement Phi for cues was not consistent, varying from .24 for Case 1 to .69 for Case 2. Totals were M = .46, SD = .19.

10. The intraclinician agreement Phi for diagnostic statements was not consistent, varying from .14 for Case 3 to .32 for Case 2. Totals were M = .26, SD = .08.

11. The intraclinician agreement Phi on cues collected was higher than the inter/intraclinician agreement Phi on cues collected for Cases 2, 3, and 4 but not for Case 1.

12. The intraclinician agreement Phi on diagnostic statements was higher than the inter/intraclinician agreement Phi on diagnostic statements made for all four cases.

13. With two exceptions, the Porter Index yielded slightly lower scores than did the Phi for both inter/intraclinician and intraclinician agreement on cues and diagnostic statements. The exceptions were the lower inter/intra Phi on cues for Case 2 and the lower intra Phi on cues for Case 1.

14. The clinicians for this study were lower in agreement (both Phi and Porter) on diagnostic statements made than on cues collected.

Process Measures

In the clinical encounter there are key behaviors that govern the interaction between a clinician and a case. The process measures deal with the behavioral dynamics of the clinical interaction in terms of the dependent, time-related variables. These variables include such time-dependent data as (1) the length of time for interaction with a case, (2) the number of cues collected and diagnostic statements made, (3) when in the session the cues are collected and diagnostic statements are made, (4) the average time that cues and statements are made in the interaction, and (5) what relationships, if any, exist between cues collected and statements of hypotheses made regarding diagnosis. The process measures thus reflect the way in which reading clinicians collect information on a case and subsequently arrive at diagnostic decisions. In terms of the Inquiry Theory, process measures deal with the strategy used by a clinician in diagnosing a case. The analysis measures for process are based on the data obtained in the observational and debriefing parts of the clinical sessions.

Results of process measures data analysis will be presented according to three major types of descriptive statistics. These were developed to answer questions relative to clinical problem-solving strategy, or the manner in which a clinician arrived at a diagnosis. The three types of statistics and the results presented are:

1. Basic process statistics, which summarize data relevant to the times that cues and statements first appeared in the clinical interaction for individual sessions and for combinations of sessions.

2. Process-agreement statistics, including correlation and partitioned Phi coefficients, which are intended to reflect clinicians' agreement on the time/order in which cues are requested and diagnostic statements made.

3. Cue-to-statement relationship statistics, which are intended to indicate the degree of relationship between cues and statements of hypotheses as suggested by the observational session

and debriefing data (e.g., whether certain cues are more frequently used to confirm certain hypotheses).

Basic Process Statistics

The basic process statistics are calculated to provide a summary of the dependent variables for each clinical session. The basic process statistics reported for this study include:

1. total elapsed time of session in minutes;

2. total number of cues collected;

3. average time that cues were collected in a given session, shown as a fraction of the total elapsed time of the session;

4. the standard deviation of the average time cues were collected in a given session;

5. total number of diagnostic statements mentioned;

6. average time that statements were made in a given session, shown as a fraction of the total elapsed time of the session; and

7. the standard deviation of the average time that statements were made in a given session.

Table 16 shows the basic process statistics data for four simulated cases in reading. The first column in the table lists the case numbers. The second column indicates the statistics (mean and true standard deviation) that are reported. Column 3 indicates the mean and true standard deviation for total elapsed time of the sessions in minutes for each of the four cases. The fourth and fifth columns show the mean and true standard deviation of the total number of cues collected and statements made, respectively, for each of

Table 16.--Basic process statistics data for four simulated cases in reading.

Case Number (1)	Statistics (2)	Total Time (3)	Total Cues (4)	Total Statements (5)	Average Cue Time (6)	Std. Dev. Cue Time (7)	Average Statement Time (8)	Std. Dev. Statement Time (9)
1 (n=6)	M =	27.33	15.17	21.67	.43	.31	.39	.30
	SD =	5.61	1.72	3.78	.05	.01	.07	.03
2 (n=6)	M =	34.17	22.17	26.50	.49	.29	.47	.28
	SD =	6.34	4.88	6.16	.05	.03	.04	.05
3 (n=6)	M =	36.50	24.83	26.00	.55	.31	.58	.31
	SD =	6.50	8.66	10.16	.09	.03	.08	.05
4 (n=6)	M =	36.17	21.67	31.17	.62	.30	.49	.31
	SD =	11.50	7.03	6.24	.05	.03	.04	.04

Case Difficulty Scale

Easiest — Most Difficult

2 4 1 3

Cases

the four cases. In columns 6 and 7, the first rows for each case indicate the mean (average) and variation (shown as standard deviation) of the average time that cues were collected in the specific sessions for each of the four cases; the second rows for each case indicate the true standard deviation for each value in the two columns. The values in the sixth column indicate the fraction of the total mean elapsed time for cues (and the true standard deviation of the mean time for cues) of the specific sessions for each case. For example, if the mean of the average cue fractional time was .50, then the mean of the average cue-collection time was exactly halfway through the session. This would indicate that, on the average, the same number of cues was collected by the clinicians for a particular case in the first half as in the last half of a session. In columns 8 and 9, the first rows for each case indicate the mean (average) and variation (shown as standard deviation) of the average time that statements were made in the specific sessions for each of the four cases; the second rows for each case indicate the true standard deviation for each mean in the two columns. The values in the eighth column indicate the fraction of the total mean elapsed time for diagnostic statements (and the true standard deviation of the mean time for diagnostic statements) of the specific sessions for each case. For example, if the mean of the average statement fractional time was .25, then the mean of the average statement collection time was one quarter of the way through the session. This would indicate that, on the average, the clinicians made more statements for a particular case in the first quarter than in the last three quarters of a session.

The average total time of the session, approximately 27 minutes, was less for Case 1 than for the other three cases, all of which had similar mean values. Case 1 also had, on the average, fewer cues collected and fewer diagnostic statements made than the other three cases. The lower mean totals for time, cues, and statements could possibly be a result of case differences. Before data collection, Case 1 had been determined to be the second most difficult of the four cases. (See Chapter III.) The means of the average cue fractional times for Case 1 and Case 2 were .43 and .49, respectively. This would indicate that, on the average, more cues were collected in the first half than in the second half of a session for Case 1 and Case 2. The means for average cue time for Case 3 and Case 4 indicate that, on the average, more cues were collected in the second half than in the first half of a session for both Case 3 and Case 4, with average cue times of .55 and .62, respectively. Case 1, the second most difficult case, had the lowest value for the four cases on average cue time. The mean of the average statement fractional time for Case 1 (.39) was the lowest value of the four cases.

Process-Agreement Statistics

One of the assumptions of the Inquiry Theory, other factors held constant, is that clinicians whose memories and strategies are held in common should have common behavior in terms of cues collected and diagnostic statements made. The process-agreement statistics are designed to reflect those behavioral agreements when several clinicians diagnose the same SIMCASE or when one clinician diagnoses an alternate form of the same SIMCASE.

Three types of statistics comprise process measurement: correlation statistics, partitioned Phi coefficients, and cue-to-statement relationship statistics.

Correlation.--The statistics that are calculated to determine the agreement between an individual and a group are based on the correlation coefficients for each session: (1) between one clinician's cue times and the average of the other clinicians' cue times and (2) between one clinician's statement times and the average of the other clinicians' statement times. The statistical measure used to determine agreement was the Pearson product-moment correlation coefficient (r_{xy}). The r to Z transformation was not used because data were based on a normal distribution. All correlation coefficients were calculated to include the times of those cues (and statements) that one clinician collected and no other clinician collected (or mentioned) for the same case, i.e., missing data.

Table 17 shows average r and standard deviation for cue times, including cues that one clinician collected and no other clinician collected (i.e., missing data) for four cases. The table indicates that the highest mean correlation for cue times occurred for Case 4, the value being .45. Case 4 was the second easiest case. The mean correlations for each of the other three cases had somewhat lower but more consistent values. The tables for each case are presented in Appendix H. (See Tables H10, H11, H12, and H13.) The data indicate that the magnitude of the relationship between the average individual cue time and average group cue time, including those cues collected by the individual but not the group, was low (M r = .36).

Table 17.--Means and standard deviations for value of r on cue times, including missing data, for four cases.

Case Number	M r	SD
1	.30	.17
2	.32	.15
3	.36	.18
4	.45	.13
	Mean total = .36	
	SD = .07	

Table 18, which is based on the data from Tables H14, H15, H16, and H17 for each case (see Appendix H) shows the average r and standard deviation for diagnostic statement times, including diagnostic statements that one clinician made and no other clinician made (i.e., missing data) on the same case. Data are presented for four cases. As the table indicates, the overall mean correlation value for diagnostic statement time was fairly consistent across all cases.

The data indicate that the magnitude of the relationship between individual statement time and group statement time, including those statements made by the individual but not by the group, was low (M r = .24). However, the direction of the relationship was positive. The highest correlation for diagnostic statement time was $M\ r = .30$ for Case 4. Looking at the correlation for diagnostic statement time (Table 18) and the correlation for cue time (Table 17), it can be seen that the individual clinicians compared with a group of clinicians diagnosing the same case showed greater magnitude in relationship for cue time than for statement time.

Table 18.--Means and standard deviations for value of *r* on diagnostic statement times, including missing data, for four cases.

Case Number	*M r*	*SD*
1	.20	.07
2	.25	.07
3	.22	.09
4	.30	.07
	Mean total = .24	
	SD = .04	

Partitioned Phi coefficients.--The partitioned Phi coefficient reflects agreement among clinicians on the time/order in which diagnostic statements made by one clinician are compared to diagnostic statements made by other clinicians on the same case; i.e., inter/intra-clinician agreement. The coefficients for partitioned Phi are calculated in a manner similar to that used in the product analysis (see Inter/Intraclinician Agreement) except that for partitioned Phi coefficients the contingency tables are calculated within four time periods or quarters instead of across the whole session. For example, if a clinician mentioned a particular diagnostic statement during the first quarter of any session, the time of that diagnostic statement was included in the calculations for the first quarter. The four partitions represent clock time into the session; i.e., Partition 1 includes the first 25 percent or first quarter of the time of the session and Partition 2 includes 50 percent or half of the time of the session.

It should be noted that the data used for time partitions were based on subjective decisions regarding the diagnostic statements the clinicians made on the H/OCL. (See Procedures in Chapter III.) The researcher identified and coded each response that was considered a diagnostic statement (the identification of a factor or variable that helps determine the state or condition of a student's reading performance) using the RDCL. Diagnostic statements for which no value judgment (strength or weakness) was made (e.g., "I was looking at instant word recognition.") were coded as observation statements. An attempt was made to represent the clinician's intent by referring to the way in which the clinician matched his own written diagnosis to the RDCL.

Table 19 shows the mean and standard deviation of Phi for four time partitions across four cases. As indicated in the table, the agreement among clinicians in the time/order that statements were made in the session was very low, showing mostly negative or no agreement. The low values appeared to be consistent across all cases for the four time partitions. In view of the previously reported low agreement of clinicians on diagnostic statements, it might not be surprising to find low partitioned Phi coefficients. However, the degree to which the clinicians showed lack of agreement on the partitioned Phi might to some extent be accounted for by the subjective coding used in analyzing the data.

Table 19.--Inter/intraclinician Phi on four time partitions for diagnostic statements and four cases.

Case Number	Partition 1		Partition 2		Partition 3		Partition 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	.01	.17	.02	.26	.01	.23	.03	.22
2	-.05	.23	-.01	.14	-.01	.20	-.02	.15
3	-.05	.20	-.01	.20	-.01	.18	-.05	.11
4	-.06	.15	.00	.21	-.00	.17	.01	.15

Cue-to-Statement Relationship Statistics

To compare the cue-to-statement relationship statistics across sessions required classification of the relationship between cues collected and diagnostic statements of hypotheses made. The classification method used for this study can be seen in column 2 (Relationship) of Table 20. The cue-to-statement relationship statistics were used to determine if there was a pattern in the relationship between the cues collected by a given clinician on a specific case and the statement of hypotheses (or other diagnostic statements) made by that same clinician on the same case.

Unlike research involving physicians, in research of reading clinicians in clinical diagnosis it is very difficult to distinguish between statements of hypotheses (i.e., statements of high utility in reaching a problem solution) and simple observational statements (i.e., statements of low or zero utility in reaching a problem solution). Whereas physicians are accustomed to dealing with a more precise and standardized vocabulary and more exact data (i.e., blood tests, urinalysis, etc.), reading clinicians lack a standardized vocabulary for diagnosis, and their diagnostic data often include informal, subjective tests. In reading, predictions concerning the effect of clinical memory and clinical strategy on the generation and use of hypotheses to direct clinical inquiry are still conjectural in terms of explication of the Inquiry Theory. Therefore, the use of early hypothesis generation to determine patterns of clinical strategy (hypothetico-deductive approach), as suggested in the explanation of the Inquiry Theory in Chapter I (see Hypothesis-Generation Corollary), are still under investigation in the field of reading. However, although this researcher did not intend to investigate in depth the use of the hypothetico-deductive approach in the diagnostic inquiry of reading clinicians, it was of interest and import to provide preliminary data in this area to facilitate future research efforts.

The method chosen in this study for initial investigation of the hypothesis-generation corollary of the Inquiry Theory in reading involved a number of subjective decisions. Those decisions included (1) selecting and coding relationship responses on the H/OCL, (2) combining responses to fit the number allowed by the computer program, and (3) classifying the relationship responses into two major categories for analysis.

The data for cue-to-statement relationship statistics represent average values on a scale of 0 to 100 for percentage of times a particular relationship appeared for cues or diagnostic statements. For example, in Table 20, Subject 103, Run 3, indicated that he had a "hunch" 12.5 percent of the time for cues on Case 1. On the other hand, Subject 107, Run 3, only indicated that he had a "hunch" 5.56 percent of the time for cues on the same case.

The relationship statistics were designed to reflect clinical strategy as defined by the Inquiry Theory. The relationship analyses used in this study concerned (1) the pattern of relationship between cues (or diagnostic statements) and hypotheses or observations made by the clinician during the clinical session and (2) the average of percentage of number of times each relationship occurred for a given group of cues (or diagnostic statements). Data for cues are presented first.

Vinsonhaler (1979b) reported that two types of strategies appear to characterize the behavior of clinicians in the clinical encounter: first, the strategy in which problem solvers tend to direct the inquiry process by the use of hypotheses about the problems of a case--i.e., cues or information is gathered to <u>test</u> the specific hypotheses; and second, the strategy in which problem solvers use cue collection to direct their inquiry--i.e., certain types of cues are collected and <u>then</u> statements (including diagnostic statements) are made. The data for the relationships of cues and diagnostic statements to statements of hypotheses were interpreted in light of these two types of strategies.

On the H/OCL, which the clinician subjects were given in the debriefing part of the clinical session, only relationship numbers 1-8 appeared as choices, and subjects were permitted to check more than one number. Later, when the program for analyzing the data was developed, it was necessary to assign additional numbers (9 through 14) in those instances in which more than one relationship had been checked. Subjects had not checked more than two relationship numbers within the two groups (relationship numbers 1-4 and 5-8) of the original eight relationship numbers.

The relationship responses on the H/OCL that were used in this research to be indicative of using a variable set of hypotheses to direct inquiry were Hunch (#1), Confirm Hunch (#5), Disconfirm Hunch (#6), Hunch and Confirmed (#9), and Hunch and Disconfirmed (#10). The relationship responses that were determined to characterize cue-directed or discovery-type inquiry included Just Wanted Information (#2), Usually Get Information (#3), Other (#4), Suggest Hunch (#7), Other (#8), Wanted Information and Confirmed (#11), Wanted Information and Suggested Hunch (#12), Usually Get Information and It Confirmed Hunch (#13), and Usually Get Information and It Suggested Hunch (#14).

Table 20 is an example, using Case 1, to show the average of percentage of number of times each relationship from the H/OCL appeared for cues collected across all statements during the clinical session. The percentages for hypothesis-directed inquiry and cue-directed inquiry should add to 100 _if_ a relationship occurred for

Table 20.--Average of percentage of number of times each relationship appeared for *cues*--Case 1.[a]

Rel. Number	Relationship	Subject 103 Run 1	Subject 103 Run 3	Subject 107 Run 1	Subject 107 Run 3	Subject 108 Run 2	Subject 102 Run 2	Row Total	M	SD
1	Hunch	0.0	12.50	2.94	5.56	3.33	11.11	35.44	5.91	4.92
2	Just Wanted Info.	0.0	6.25	0.0	16.67	23.33	8.33	54.58	9.10	9.32
3	Usually Get Info.	26.56	21.88	36.27	27.22	15.00	18.65	145.58	24.26	7.50
4	Other	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	Confirm Hunch	6.25	0.0	34.31	35.56	6.67	28.97	111.76	18.63	16.02
6	Disconfirm Hunch	0.0	0.0	5.88	3.33	3.33	0.0	12.54	2.09	2.47
7	Suggest Hunch	23.44	40.63	0.0	5.00	17.22	6.55	92.84	15.47	15.02
8	Other	0.0	0.0	0.0	0.0	11.11	0.0	11.11	1.85	4.54
9(1+5)	Hunch & Confirmed	0.0	0.0	8.82	0.0	13.33	18.06	40.21	6.70	7.91
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	0.0	6.25	5.88	6.67	0.0	0.0	18.80	3.13	3.44
12(2+7)	Info. & Suggested Hunch	6.25	0.0	0.0	0.0	0.0	0.0	6.25	1.04	2.55
13(3+5)	Usually Get & Confirmed Hunch	0.0	0.0	5.88	0.0	0.0	0.0	5.88	.98	2.40
14(3+7)	Usually Get & Suggested Hunch	6.25	6.25	0.0	0.0	0.0	0.0	12.50	2.08	3.23

[a]Based on scale of 0 to 100.

each cue and each diagnostic statement mentioned. However, in this study there was not always a one-to-one match or relationship between cues and hypotheses (or other statements). In Tables 21 and 23, those percentages out of 100 for which relationships occurred are shown in the last column of each table. A look at Table 20, by relationships, indicates that for Case 1 the most frequently used relationship for cues was relationship number 3, Usually Get Information. The second most frequently used relationship was number 5, Confirm Hunch.

Looking at the data in the table by subject, it appears that when subjects were asked why they collected certain information (relationship numbers 1-4), in general, most of them indicated that they either just wanted the information or usually got the information for a case they were diagnosing. This would seem to indicate that, in general, the subjects diagnosing Case 1 said that they used a strategy of cue collection in which certain cues were collected and then statements made (including diagnostic statements) concerning the information provided by the cues. When asked what the information told them (relationship numbers 5-8), in general, most of the clinicians diagnosing Case 1 responded that the information confirmed a hunch or suggested a hunch. Tables for Cases 2, 3, and 4 appear in Appendix H. (See Tables H18, H19, and H20).

By combining relationships into two categories of inquiry, it can more clearly be shown how the individual subjects responded to questions on the two types of strategies used to characterize the clinical inquiry of the eight clinicians in this study. Table 21 is a summary by subjects across all runs for all cases, using means and standard deviations, for average of percentage of number of times relationships appeared for cues in hypothesis-directed inquiry and cue-directed inquiry responses. The responses for relationship numbers 1, 5, 6, 9, and 10 were used to characterize hypothesis-directed inquiry appearing for cues. The responses for relationship numbers 2, 3, 4, 7, 8, 11, 12, 13, and 14 were used to characterize cue-directed inquiry appearing for cues.

Looking at Table 21 for the average of percentage of number of times relationships appeared for cues, Subject 101 indicated the greatest use of hypothesis-directed inquiry. However, Subject 105 indicated the second highest use of hypothesis-directed inquiry and was somewhat more consistent across all three session runs. In the same table, the mean cue-directed inquiry responses for the average of percentage of number of times relationships appeared for cues was highest for Subject 103, indicating an information-gathering approach in clinical diagnosis. The other subjects seemed to indicate use of combinations of hypothesis-directed and cue-directed inquiry on relationships appearing for cues.

Case differences for the two types of strategies can be shown by finding the overall average by case on responses for (1) hypothesis-directed inquiry items and (2) cue-directed inquiry items. Using a scale of 0 to 100, the average or mean for each subject on each case was computed for the hypothesis-directed inquiry responses and the cue-directed inquiry responses. Table 22 shows means and standard deviations across all subjects and cases for average of percentage

Table 21.--Average of percentage of number of times relationships appeared for cues using hypothesis-directed and cue-directed inquiry responses across all subjects and all cases.[a]

Subject Number	Case Number[b]	Run Number	Hypothesis-Directed Inquiry	Average for 3 Sessions & (SD)	Cue-Directed Inquiry	Average for 3 Sessions & (SD)	% in Relationship
101	3	1	58.33	67.70	33.33	28.07	91.66
101	3'	3	68.11	(9.18)	27.54	(5.02)	95.65
101	4	2	76.67		23.33		100.00
102	2	1	42.59	47.47	31.47	41.11	74.06
102	2'	3	41.67	(9.25)	58.33	(14.95)	100.00
102	1	2	58.14		33.53		91.67
103	1	1	6.25	17.79	62.50	67.15	68.75
103	1'	3	12.50	(14.91)	81.26	(12.45)	93.76
103	3	2	34.62		57.70		92.32
104	4	1	51.66	43.17	28.33	50.16	79.99
104	4'	3	50.59	(13.78)	49.42	(22.20)	100.01
104	2	2	27.27		72.72		99.99
105	3	1	53.45	56.04	29.31	34.57	82.76
105	3'	3	54.83	(3.37)	37.28	(4.56)	92.11
105	4	2	59.85		37.13		96.98
106	4	1	35.26	38.41	60.89	57.13	96.15
106	4'	3	48.96	(9.38)	41.50	(14.13)	90.46
106	2	2	31.02		68.99		100.01
107	1	1	51.95	47.34	48.03	49.48	99.98
107	1'	3	44.45	(4.03)	55.56	(5.50)	100.01
107	3	2	45.63		44.84		90.47
108	2	1	31.25	35.35	60.42	58.41	91.67
108	2'	3	48.14	(11.31)	48.15	(9.42)	96.29
108	1	2	26.66		66.66		93.32

[a]Based on the average of percentage of number of times each relationship appeared for cues. (See Table 20 and Tables H18, H19, and H20 in Appendix H.)

[b]1', 2', etc. = alternate forms of Case 1, Case 2, etc.

of number of times relationships for hypothesis-directed and cue-directed inquiry responses appeared for cues. (See Row Total column for Table 20.)

Table 22.--Means and standard deviations across all subjects and cases for average of percentage of number of times relationships for hypothesis-directed and cue-directed inquiry responses appeared for cues.

Case Number	Hypothesis-Directed Inquiry		Cue-Directed Inquiry	
	M	*SD*	*M*	*SD*
1 (*n*=6)	33.33	21.44	57.92	16.37
2 (*n*=6)	36.99	8.25	56.68	15.06
3 (*n*=6)	52.50	11.41	38.33	11.33
4 (*n*=6)	53.83	13.72	40.10	13.78
	Mean total = 44.16		Mean total = 48.26	
	SD = 10.52		*SD* = 10.48	

Data in Table 22 suggest that when subject responses using relationships characterizing hypothesis-directed inquiry for cues were compared by case to subject responses using relationships characterizing cue-directed inquiry for cues, subjects diagnosing Cases 1 and 2, on the average, indicated by their responses that they used more cue-directed inquiry than hypothesis-directed inquiry. This would imply that they collected cues based on a fixed strategy. Conversely, the subjects for Cases 3 and 4, on the average, indicated by their responses that they used hypothesis-directed inquiry, implying that they collected cues based on a variable strategy; i.e., their hypotheses (hunches) determined the information they collected. (See Appendix H.)

Table 23 shows the average of percentage of number of times each relationship from the H/OCL appeared for diagnostic statements made for Case 1. A look at the table indicates that the most frequently used relationship for statements on Case 1 was relationship number 5, Confirm Hunch.

As mentioned earlier, by combining relationships into categories of hypothesis-directed and cue-directed inquiry, it can more clearly be shown how the individual subjects responded to questions formulating the two types of inquiry. Table 24 is a summary for average of percentage of number of times relationships appeared for diagnostic statements using hypothesis-directed and cue-directed inquiry responses.

The same relationship responses (numbers 1, 5, 6, 9, and 10) that were used to characterize hypothesis-directed inquiry appearing for cues also apply to diagnostic statements. Likewise, the same relationship responses (numbers 2, 3, 4, 7, 8, 11, 12, 13, and 14) that characterized cue-directed strategy appearing for cues apply to diagnostic statements. Means and standard deviations were computed using the same procedures appearing for the previously described cue data. The tables for statement computations include Table 23 and Tables H21, H22, and H23 in Appendix H. As shown by data in Table 23, inconsistencies in subject-reported strategy become apparent, e.g., Subjects 103 and 108.

Table 23.--Average of percentage of number of times each relationship appeared for diagnostic statements--Case 1.[a]

Rel. Number	Relationship	Subject 103 Run 1	Subject 103 Run 3	Subject 107 Run 1	Subject 107 Run 3	Subject 108 Run 2	Subject 102 Run 2	Row Total	M	SD
1	Hunch	0.0	15.63	4.35	7.69	2.00	9.52	39.19	6.53	5.67
2	Just Wanted Info.	0.0	12.50	0.0	8.97	16.00	1.59	39.06	6.51	6.94
3	Usually Get Info.	44.74	18.75	39.13	25.64	24.00	35.71	187.97	31.33	10.04
4	Other	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	Confirm Hunch	15.79	0.0	39.13	35.58	6.00	25.40	121.90	20.32	15.32
6	Disconfirm Hunch	0.0	0.0	6.52	11.54	2.00	0.0	20.06	3.34	4.75
7	Suggest Hunch	35.09	48.96	0.0	6.73	26.00	11.11	27.89	21.32	18.69
8	Other	0.0	0.0	0.0	0.0	14.00	0.0	14.00	2.33	5.72
9(1+5)	Hunch & Confirmed	0.0	0.0	4.35	0.0	6.00	16.67	27.02	4.50	6.50
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	0.0	2.08	2.17	3.85	0.0	0.0	8.1	1.35	1.61
12(2+7)	Info. & Suggested Hunch	2.63	0.0	0.0	0.0	0.0	0.0	2.63	.44	1.07
13(3+5)	Usually Get & Confirmed Hunch	0.0	0.0	4.35	0.0	0.0	0.0	4.35	.725	1.78
14(3+7)	Usually Get & Suggested Hunch	1.75	2.08	0.0	0.0	0.0	0.0	3.83	.638	.994

[a]Based on scale of 0 to 100.

Table 24.--Average of percentage of number of times relationships appeared for diagnostic statements using hypothesis-directed and cue-directed inquiry responses across all subjects and all cases.[a]

Subject Number	Case Number	Run Number	Hypothesis-Directed Inquiry	Average for 3 Sessions & (SD)	Cue-Directed Inquiry	Average for 3 Sessions & (SD)	% in Relationship
101	3	1	78.57	73.66	21.43	25.06	100.00
101	3'	3	76.37	(6.70)	23.63	(4.52)	100.00
101	4	2	66.03		30.13		96.16
102	2	1	71.16	58.28	28.51	41.50	96.67
102	2'	3	52.10	(11.15)	47.90	(11.26)	100.00
102	1	2	51.59		48.11		99.70
103	1	1	15.79	26.28	84.21	72.53	100.00
103	1'	3	15.63	(18.30)	84.37	(20.37)	100.00
103	3	2	47.41		49.01		96.42
104	4	1	44.93	33.09	55.07	66.09	100.00
104	4'	3	34.35	(12.51)	63.21	(12.71)	97.56
104	2	2	20.00		80.00		100.00
105	3	1	73.44	59.64	26.57	34.12	99.67
105	3'	3	52.35	(12.00)	28.90	(11.11)	81.25
105	4	2	53.13		46.88		100.01
106	4	1	31.31	34.66	68.69	65.34	100.00
106	4'	3	43.12	(7.38)	56.87	(7.39)	99.99
106	2	2	29.55		70.45		100.00
107	1	1	54.35	51.98	45.65	44.80	100.00
107	1'	3	54.81	(4.51)	45.19	(1.10)	100.00
107	3	2	46.77		43.55		90.32
108	2	1	27.68	30.40	72.32	68.27	100.00
108	2'	3	47.52	(15.94)	52.48	(14.20)	100.00
108	1	2	16.00		80.00		96.00

[a]Based on the average of percentage of number of times each relationship appeared for statements. (See Table 23 and Tables H21, H22, and H23 in Appendix H.)

[b]1', 2', etc. = alternate forms of Case 1, Case 2, etc.

Case differences for the two types of strategies can be shown by finding the overall average by case on responses for hypothesis-directed and cue-directed inquiry. Table 25 shows means and standard deviations across all subjects and cases for average of percentage of number of times relationships for hypothesis-directed and cue-directed inquiry responses appeared for diagnostic statements. (See Row Total column for Table 23.)

Table 25.--Means and standard deviations across all subjects and cases for average of percentage of number of times relationships for hypothesis-directed and cue-directed inquiry responses appeared for diagnostic statements.

Case Number	Hypothesis-Directed Inquiry		Cue-Directed Inquiry	
	M	*SD*	*M*	*SD*
1 (*n*=6)	34.70	20.72	64.59	20.10
2 (*n*=6)	41.34	19.10	58.61	19.20
3 (*n*=6)	63.43	14.70	32.18	11.34
4 (*n*=6)	45.48	12.74	53.48	13.63
	Mean total = 46.25		Mean total = 52.22	
	SD = 12.32		*SD* = 14.11	

Data in the table suggest that when subject responses using relationships characterizing hypothesis-directed inquiry for diagnostic statements are compared by case to subject responses using relationships characterizing cue-directed inquiry for diagnostic statements, subjects diagnosing Case 3, the most difficult case, on the average indicated by their responses that they used hypothesis-directed inquiry to a greater extent than did subjects for the other cases. It also appears that Cases 1, 2, and 4 showed more responses indicating use of cue-directed than hypothesis-directed inquiry. Overall, the subjects indicated more use of information-gathering strategy, or cue-directed inquiry, than hypothesis-directed inquiry in making diagnostic statements about the cases.

Process Measures: Summary

The process measures deal with clinical problem-solving strategy, or the manner in which a clinician arrives at a diagnosis. The behavioral dynamics of the clinical interaction, which includes the clinician's problem-solving strategy, are measured statistically in terms of the dependent, time-related variables, i.e., length of time for case interaction, number of cues collected, and number of diagnostic statements made.

Results of the formal process data analysis are presented according to three major types of descriptive statistics: (1) basic process statistics, which summarize data relevant to times cues and statements first appeared in the clinical interaction for individual and combinations of sessions; (2) process-agreement statistics, which include correlation and partitioned Phi coefficients, and are intended to reflect agreement among clinicians on the time/order in which cues are collected and diagnostic statements are made; and (3) cue-to-statement relationship statistics, which are intended to indicate the degree of relationship between cues and statements of hypotheses as suggested by the H/OCL (debriefing) data.

The process-measures data seem to suggest the following:

1. Case 1, the second most difficult case, had the lowest average total time of the clinical sessions--an average of approximately 27 minutes total elapsed time.

2. Case 1 had, on the average, the fewest cues collected; average cues totaled 15.

3. Case 1 had, on the average, the fewest diagnostic statements made; average statements totaled 22.

4. Case 1 had, on the average, the lowest average cue fractional time; more cues were collected in the first half than in the second half of a session. The mean of the average cue time for Case 1 was $\underline{M}$ = .43, $\underline{SD}$ = .05.

5. Case 1 had, on the average, the lowest average statement fractional time. The mean of the average statement time for Case 1 was $\underline{M}$ = .39, $\underline{SD}$ = .07.

6. Case 2, the easiest of the four cases, had, on the average, more cues collected in the first half than in the second half of a session. The mean average cue time for Case 2 was $\underline{M}$ = .49, $\underline{SD}$ = .05.

7. Case 3, the most difficult case, and Case 4, the second easiest case, had, on the average, more cues collected in the second half than in the first half of the session. The mean average cue time for Case 3 was $\underline{M}$ = .55, $\underline{SD}$ = .09; and for Case 4 $\underline{M}$ = .62, $\underline{SD}$ = .05.

8. The highest mean correlation for cue times occurred for Case 4, the second easiest case, at $\underline{M}\ \underline{r}$ = .45, $\underline{SD}$ = .13, with the other three cases having somewhat lower but more consistent mean correlations.

9. The overall mean correlations for statement time were fairly consistent, with the highest value for Case 4 at $\underline{M}$ $\underline{r}$ = .30, $\underline{SD}$ = .07.

10. The average individual clinician compared with a group of clinicians diagnosing the same case showed greater magnitude in relationship for cue time ($\underline{M}$ $\underline{r}$ = .36, $\underline{SD}$ = .07) than for diagnostic statement time ($\underline{M}$ $\underline{r}$ = .24, $\underline{SD}$ = .04).

11. The agreement among clinicians on the time/order in which diagnostic statements were made by one clinician when compared to other clinicians on the same case (inter/intraclinician agreement), as reflected by the partitioned Phi coefficients for statements, indicated very low, mostly negative, or no agreement; the range of the means was -.003 (Case 4) to .03 (Case 1).

12. In general, the subjects diagnosing the four cases indicated by their responses concerning <u>why</u> they asked for certain information that they used some combination of hypothesis-directed and cue-directed inquiry when collecting information on a reading case. The mean total for the average of percentage of number of times relationships appeared for cues was $\underline{M}$ = 44.16, $\underline{SD}$ = 10.52 for hypothesis-directed inquiry and $\underline{M}$ = 48.26, $\underline{SD}$ = 10.48 for cue-directed inquiry. (See Table 22.)

13. Subject 101, Run 2, Case 4, with the highest average of percentage of number of times relationships appeared for cues (76.67 percent), indicated that he used mostly a hypothesis-directed approach or strategy in making decisions about that case. (See Table 21.)

14. In general, the clinicians in this study indicated by their responses to the question of _why_ they asked for certain information on a reading case that they used primarily cue-directed inquiry in making diagnostic statements about the cases. The mean total for average of percentage of number of times relationships appeared for diagnostic statements was $\underline{M} = 46.25$, $\underline{SD} = 12.32$ for hypothesis-directed inquiry and $\underline{M} = 52.22$, $\underline{SD} = 14.11$ for cue-directed inquiry. (See Table 25.)

15. Subject 101, Run 1, Case 3, with the highest average of percentage of number of times relationships appeared for statements (78.6 percent), indicated that he used mostly a hypothesis-directed strategy in diagnosing that case. (See Table 24.)

The Informal Product/Process Measures

It may be recalled that the second principle of the inquiry theory describes those factors that govern the clinician's behavior during the clinical encounter. This second principle states that those events or behaviors that occur in the clinical interaction are determined probabilistically by the case and the clinician's memory and strategy.

The purpose of the informal product/process measurement was to gain further insights into the clinical behavior of reading clinicians, the clinical behavior being defined in terms of the data base represented in clinical memory and the diagnostic routines represented in clinical strategy. Additionally, it was important to attempt to determine the source (schema) that allowed the clinician's data base (e.g., cue collection) to be translated into an action (e.g., hypothesis generation).

It was recognized that extended questioning in the debriefing might overburden the clinician subjects, but at the same time it was desirable to investigate their clinical/diagnostic behavior through measures of informal self-reporting. Therefore, the informal product/process portion of the third session was limited to six questions. One of the questions was designed to elicit the clinicians' ideas about the content of a diagnosis or the data base (memory). Another question was designed to elicit information on the way in which the diagnosis was conducted, that is, the routine used (strategy). Four questions were designed to generate responses about the source of the data base or that which allows the clinician to generate a diagnosis. The six questions, in the form used by the examiner when listening to tapes of the clinicians' responses, appear in Appendix C. These questions were generated from a diagnostic model of reading and learning (Sherman et al., 1978). The informal questioning followed the last session for each clinician in the study and was used to guide his thinking about the third case with which he worked. The paraphrased responses of representative clinicians (104 & 106) to the questions in the informal portion of the session appear in Tables 26-28. The complete, unedited dialogue for Subject 104 appears in Appendix D. Also in Appendix D are the paraphrased responses of the other clinicians in the study.

In looking at the data base (memory) reported by the clinician, the interest was in determining <u>what</u> were the descriptors of the data base (i.e., What information should be included in a "good" diagnosis?).

In the diagnostic routine, the concern was with the task of information gathering or how the data base (memory) was translated into action (strategy) (i.e., How do you usually go about a diagnosis?).

The third important aspect of the informal product/process measures involved hypothesis generation or the source (schema) of the data base that allowed memory to be translated into strategy (i.e., How did you know . . . that the reading fluency was low?).

Table 26 shows one sample of responses to informal questions regarding memory and strategy. The sample reflects one kind of material that clinicians said was contained in the data base (Question 1) and the diagnostic routine. Sample 1, Question 1, appears to reflect a straightforward type of data collection, saying, in effect, "I do this in order to find out this and then I do this"--a cue-directed process. Table 27 shows another sample of responses to the same questions as Sample 1. However, the second sample is more reflective of data collection for the purpose of making a comparison: "Depending on what I found I would do this or this"--a hypothesis-directed process. Interestingly, some of the subjects answered the question about what should be included in a good diagnosis in behavioral or process terms of how rather than in content or product terms. Others used a combination of behavioral and content-type responses. This latter observation also held true for the second question regarding the "how" of diagnosis. It appears that, based

Table 26.--Sample 1 of responses to informal questions relating to memory (1) and strategy (2)--Subject 104.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. Unusual factors such as: Physical (sight & hearing) Emotional (child abuse) Environmental (abnormal)	1. Informal tests
2. What kid knows about words	2. Sight words
3. Informal oral reading, usually child's choice of material, to get hunch of "where he is"	3. Phonics
4. Slosson Oral Reading Test to get grade score	4. Structural analysis
5. Ekwall Inventory for miscues on words in isolation to determine understanding of phonics system	5. Application
6. More formal phonics test to see if he knows initial and final consonants, blends, etc.	6. Comprehension (limited assessment)
7. Instant sight-word recognition	
8. Oral reading to determine if word-by-word or phrase reader (fluency)	
9. Listening comprehension	

Table 27.--Sample 2 of responses to informal questions relating to memory (1) and strategy (2)--Subject 106.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. The Wechsler or some kind of an intellectual assessment of their potential strengths and weaknesses.	1. I would start with family background and school background.
2. Comparison of oral and silent reading to find out what kinds of decoding problems the child is having as well as the oral and silent comprehension.	2. I would talk with the teacher.
3. Listening comprehension to compare with the WISC.	3. I would check the physical for hearing and if a problem I would pursue the auditory.
4. Word recognition test to compare with kinds of errors made on the Durrell oral reading.	4. Then I would take care of other things I mentioned in a "good" diagnosis.
5. Word analysis to compare with a Dolch and to get an indication as to whether or not they are sight-word readers or have analysis or decoding skills.	
6. Definitely a visual and hearing screening. (I give these early to eliminate physical problems.)	
7. Family and school history to understand whole child and find out if there are physical problems relating to reading difficulty or if familiar kinds of problems affecting attitude and motivation; also to check excessive absences.	
8. Then depending on what I found, I would go to more specific things like checking auditory blending and discrimination and check digit span on WISC and if he didn't get it, I would give auditory memory and visual memory from Durrell.	
9. Child's comments in an interview are also important to find out how he feels about reading and what he thinks his reading problems are.	

Table 28.--Sample of responses to four questions regarding specific aspects of reading--Subject 104, Run 3, Case 4.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other Response: Slosson scores--grade 2; he's in fourth grade.	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other Response: Confused hodge-podge understanding of phonics. Gets initial but not final consonant. Miscues.
3a. Was reading comprehension low?	**3b. How do you know?**	**4a. Was reading fluency low?**	**4b. How do you know?**
✓ Yes — Oral & Silent ✓ No — Listening	✓ Model __ Experience __ Authority __ Other Response: Can't apply systems, doesn't know words or phonics but can comprehend fifth grade listening material. Scores: Silent and oral two years lower than listening comprehension.	✓ Yes __ No	✓ Model __ Experience __ Authority __ Other Response: Word-by-word plodder. Doesn't understand system so can't expect him to apply it.

on the informal data, clinicians in this study in general did not share a clinical memory, in terms of a data base, i.e., information that should be included in a "good" diagnosis. Likewise, it appears that these same clinicians did not share a clinical strategy in terms of a diagnostic routine, i.e., how to go about a diagnosis.

Table 29 is a tabulation of responses by eight clinicians to specific questions regarding reading skills for four cases, two clinicians per case. The results suggest that all subjects for all cases employed a schema in making determinations about specific skill areas of reading. This schema appears to reflect some theoretical process based on cause/effect relationships; authority, based on test data or teacher report, etc.; and other, when it was not based on process, experience, or authority. No responses were coded "experience."

It appears that although any one or all of the clinicians may as a normal practice use theoretic process as a source of their data base or memory, the subjects' responses did not always reflect that assumption. Therefore, it could be that some of the clinicians in this study did not consistently employ a standard or model for diagnosis, or it could be that the questions themselves failed to elicit the kinds of responses that would reflect the use of a formal theoretic process model (i.e., cause/effect relationship).

Also, it can be seen in Table 29 that clinicians diagnosing the same case were not always in complete agreement with each other

Table 29.--Results of responses by eight clinicians to specific questions regarding four areas of reading for four cases.

Case No.	Subj. No.	1a. Was instant word recognition low?		1b. How did you know?	2a. Were word analysis skills low?		2b. How did you know?	3a. Was reading comprehension low?		3b. How did you know?	4a. Was reading fluency low?		4b. How did you know?
		Yes	No		Yes	No		Yes	No		Yes	No	
1	103	✓		Authority		✓	Authority		✓	Authority	✓		Model
1	107	✓		Authority	✓		Other		✓	Model	✓		Other
2	102		✓	Model	✓		Other	✓	✓	Model	✓	✓	Other
2	108	✓		Authority	✓	✓	Other	✓	✓	Model	✓		Other
3	101		✓	Model	✓		Model		✓	Model	✓		Model
3	105	✓*		Authority	✓*		Authority	✓		Authority	✓		Model
4	104	✓		Authority	✓		Other	✓	✓	Model	✓		Model
4	106	✓		Authority	✓		Other	✓		Model	✓		Model

*Response to question was "probably."

even when asked about the same specific reading skills. The individual responses of the clinicians shed light on why it might appear that reading clinicians disagree if asked to answer "Yes" or "No" to diagnostic questions. For example, Subject 101's response to the question, "Was instant word recognition low?" was: "Not that low. He's in 7th grade and his score was 6.8 on instant word recognition. But the school district has the reputation for having children doing well and so many of his peers are probably above him. He's conceiving himself as being poor, so in that case he's probably low." (See Table D11, Appendix D.) This response might be marked "No" for low instant word recognition and left untreated. However, considering environmental factors, it might be wise to supplement what could be a strength for that student.

Informal Product/Process Measures: Summary

The informal product/process measures were designed to reflect in a less formal way than was reflected in the statistical measures the second principle of the Inquiry Theory: the probabilistic determination of behavioral interaction between a case and the clinician, represented by the clinician's memory and strategy. Six questions based on Sherman's Model of Reading and Learning (1978) were asked orally to the eight clinicians in this study following the third or last clinic session for each clinician. The questions were designed to (1) elicit information regarding the clinician's conceptions of the content or data base of a "good" diagnosis; (2) determine how the data base was translated into action, i.e., the diagnostic routine; and (3) identify the source or schema that allowed the data base (memory) to be translated into action (strategy).

In general, the clinicians' responses to the informal product/ process questions seemed to indicate that the "expert" clinicians in this study did not share a clinical memory in terms of a data base, i.e., information that should be included in a "good" diagnosis. Likewise, the responses of those same clinicians seemed to indicate that they did not share a clinical strategy in terms of a diagnostic routine or how they went about a diagnosis. Additionally, results of the clinicians' responses to informal questions suggested that all subjects employed a schema in making determinations about four areas of reading diagnosis: instant word recognition, word analysis, reading comprehension, and reading fluency. This schema appeared to reflect theoretical process models based on cause/effect relationships; authority models based on specific test data, teacher report, etc.; and some other model, reflecting a guess, an intuition, or some combination of nebulous factors.

The informal data do seem, in general, to support the statistical findings and additionally provide insights into the interrelationships among diagnostic reading factors, which make the application of statistical analysis to clinical diagnosis of reading a very difficult though seemingly justifiable pursuit.

Summary

The analysis of the data was presented in three parts: (1) formal product measures, (2) formal process measures, and (3) informal product/process measures.

The major finding indicated by the product or outcome measures was that there was little agreement among experienced, highly trained reading clinicians, using simulated reading cases, on the data they collected and the diagnostic statements they made for specific reading problems. This finding was supported by an analysis of the results of several statistical measures, including proportional agreement, commonality scores, inter/intra and intraclinician agreement Phi coefficient, and the Porter statistic.

The major finding suggested by the process measures was that the experienced, highly trained reading clinicians, using simulated reading cases, lacked an extensive and systematic method of collecting data and evaluating information about specific reading problems. This finding was supported by an analysis of the results of several statistical measures, including basic statistics (time of cue collection, number of cues collected and diagnostic statements made), correlation, partitioned Phi coefficient, and cue-to-statement relationship statistics.

The major conclusions suggested by the informal product/process data were that (1) although experienced, highly trained reading clinicians, using simulated reading cases, may as a normal practice use theoretic process as a source of their data base in clinical diagnosis, they do not consistently employ a standard or model for diagnosis and (2) although experienced, highly trained reading clinicians, using simulated reading cases, may have their own individual way of conducting a diagnosis, these reading clinicians do <u>not</u> use a common systematic or comprehensive diagnostic routine or clinical strategy when diagnosing specific reading problems.

CHAPTER V

SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Summary

Introduction

The response of reading researchers to the challenge of more effective and efficient reading diagnosis and treatment has shown little effort to investigate the way in which reading clinicians think about their students' reading problems (Shulman & Elstein, 1975).

If one assumes that correct diagnosis is a prerequisite of effective remediation, then it would seem to follow that one approach to improving remedial practices would be to improve the diagnostic skills and training of reading clinicians (Hoffmeyer & Bader, 1978). Recent studies in medicine have demonstrated how clinicians might be taught to seek answers to pertinent questions and thereby improve their clinical skills and subsequently their diagnostic and therapeutic competence (Barrows et al., 1976; Elstein, Shulman, Sprafka et al., 1978; Vinsonhaler, Wagner, & Elstein, 1977). Similar research is needed in the field of reading. The need is threefold: First, there is practical value in having a well-understood theoretic base, such as the one developed in medicine, for research on clinical problem solving in reading; second, there is efficacy in using scientifically based methods to examine clinical problem solving in reading; and third, there is augmentative value when research studies share a common methodology (Vinsonhaler, 1979a).

To move in the direction of improving problem solving among reading clinicians, the first step might be the observation and study of the decision-making and problem-solving behavior of "expert" reading clinicians for the purpose of distinguishing those reasoning skills that characterize them as "experts." Then, as in medical studies, these behaviors might be used as models in teaching effective and efficient clinical diagnosis in reading.

The Problem

The general purpose of this study was to use the Inquiry Theory of Clinical Problem Solving as a theoretic base in determining _how_ eight experienced reading clinicians diagnosed specific reading problem cases (process), and _what_ information was used in reaching a diagnosis (product or outcome). Specifically, the purpose of this study was formally to test three basic components of the Inquiry Theory. These components were (1) the agreement of reading clinicians in collecting data (cues) on a specific case in order to diagnose a reading problem, (2) the agreement of these same clinicians in making diagnostic statements, and (3) the reading clinicians' use of hypotheses to direct their diagnostic clinical inquiry. Additional informal assessment was made regarding (1) the clinicians' agreement on the information that should be included in a "good" diagnosis, (2) the clinicians' agreement on how to go about a diagnosis, and (3) the schema employed by the clinicians in making diagnostic determinations about specific skill areas of reading for a particular case.

Review of the Literature

The Inquiry Theory was developed by a team of researchers at Michigan State University to provide a formal theoretic structure that would integrate and account for the numerous concepts and empirical findings on clinical problem solving. The theory postulates that the clinical encounter involves a case and a clinician, and is characterized by the interaction that occurs between the case and the clinician's memory (problem, cue, cue value, diagnosis, treatment, and the relations between them) and strategy (information-gathering and information-processing tasks that translate memory into action).

Although the study of how information is processed has been confused and complicated by conceptually unclear terminology, research procedures in problem solving have shared a common relationship in tasks relating to realistic environments. Such procedures include, among others, (1) total-task or process-tracing studies, which are concerned with the sequential character of gathering information in order to make a decision or judgment; (2) in-basket studies, in which the decision maker receives information "input" for making a decision; and (3) tab-item methods, which provide objective, reliable data through predesignation or choices of action, such as troubleshooting electronic equipment performance failure. Of major import to investigators using these information approaches is the consideration of human behavior in actual task environments and the observation of the process of thinking and judgment in these environments. In the investigation of educational problems, studies dealing with thinking, human judgment, and decision making are lacking. A major area in education that lends itself to the study of thinking, human judgment, and decision making is reading (Shulman & Elstein, 1975). However, as in other areas involving the processing of information, the field of reading likewise has its inherent problems. Some of the concerns related to problem solving in the diagnostic process of reading include (1) the etiological and therapeutic aspects and procedural steps in diagnosis; (2) the causal debate, singular versus pluralistic and etiological versus therapeutic; and (3) the lack of standardized terms. These and other concerns are important because disagreements over such issues have complicated communication both within the reading profession and between fields relating to reading, such as learning disability and psychology.

One attempt to explicate clinical diagnosis in medicine and subsequently in reading has been the research relating to the Inquiry Theory of Clinical Problem Solving. A number of interrelated studies have been undertaken in the research on clinical diagnosis in reading. Results of these studies suggest the need and provide the impetus for the improvement of instruction, performance, and evaluation of reading clinicians.

Design and Procedures

The major objective of this study was to determine the nature and extent of agreement of eight experienced reading clinicians in terms of (1) the data they collected on specific cases in reading in order to make diagnostic decisions about those cases, (2) the diagnostic statements made for specific cases, and (3) the use of hypothesis generation in order to reach a diagnosis on specific reading problem cases.

The eight subjects in this research were chosen by university faculty recommendation from a list of reading clinicians who had taught the summer institute courses in reading diagnosis and remediation offered by Michigan State University. From among those recommended, the eight highly trained and experienced clinicians who volunteered included two male and six female clinicians. All of the clinicians were paid at a professional rate for each of the three approximately three-hour sessions (spaced no less than one week nor more than four weeks apart) in which they agreed to participate.

The clinician subjects, taken one subject at a time, were asked to obtain materials from a SIMCASE (a SIMulated CASE or set of data representing a child with a reading problem). SIMCASE materials were contained in a box, and the examiner provided the information by handing the material to the subject as the subject requested it. The clinician did not know specifically what material was in the SIMCASE; however, he was told that the information could be provided in five forms as test scores, examiner's comments, test booklets, audio recordings, and test directions. Examples were shown for each. The clinician subject was given 45 minutes in which to collect and study the SIMCASE materials. The subject was free to take notes and to retain all items of information requested.

Following the SIMCASE interaction, the clinician subject was given 25 minutes in which to write a diagnosis and an additional 25 minutes in which to write a remediation. After a short break, the subject was asked to (1) transfer his written diagnosis to the Reading Diagnostic Check List (RDCL) and (2) indicate if each of the reading factor statements in his written diagnosis was a strength, a weakness, or just an observation. Next the subject was asked to complete the Hypothesis/Observation Check List (H/OCL) by responding to the questions of <u>why</u> he had asked for each piece of information (cue) and <u>what</u> the information had told him. The subject was asked to write a brief explanation for each response to the <u>what</u> question.

The above procedures were followed for each of the three clinical sessions. Additional procedures for the third or last session only included questions designed to (1) elicit information regarding the clinician's conception of the content or data base of a "good" diagnosis, (2) determine the clinician's diagnostic routine, and (3) determine the source or schema that allowed the data base (memory) to be translated into action (strategy); i.e., there were questions dealing with four areas of reading and how they could be defined for the third or last case each subject diagnosed. The four areas were (1) instant word recognition, (2) word analysis, (3) reading fluency, and (4) reading comprehension (Sherman, 1979).

All stimulus materials used in the study, including SIMCASEs and equivalent forms, were subjected to counterbalancing to minimize systematic effects. Subjects were randomly assigned to the case order.

Subject reliability on the use of the RDCL, a previously unused instrument, was measured using a test-retest procedure. The Pearson product-moment correlation coefficient ($\underline{r}$) was calculated as the measure of relationship. The value of $\underline{r}$ ranged from .15 to .69. The Porter statistic, $\frac{A}{A+B+C}$ (see Wilcox, 1977), an index of the proportion of agreement that excludes clinician agreement not to select a cue or make a diagnostic statement, was also calculated on the same data. The Porter values ranged from .10 to .54.

The RDCL of reading factors was designed to objectify the data collection and analysis. Clinicians were asked to convert their written diagnoses to the check list, indicating strengths, weaknesses, or observations for diagnostic statements on the list corresponding to their written statements.

Analysis Measures and Results

Because of the limited sample size used in this study and because the methodology and statistical measures are still open to speculation, all findings and conclusions offered in this dissertation must be considered tentative and should not be generalized beyond this study, pending further evidence.

The analysis of the data was presented in three parts: (1) formal product measures (including proportional agreement, commonality scores, inter/intra-and intraclinician agreement Phi correlation, and the Porter statistic), (2) formal process measures (using correlation, partitioned Phi coefficients, and cue-to-statement relationship statistics, and (3) informal product/process measures (using Sherman's Model of Reading and Learning to Read).

The major findings related to product or outcome of the clinical interaction between a reading clinician and a case (SIMCASE) include the following:

1. Twenty-three percent of the same cues from the total cues available on the four cases were collected by three or more (P.A. = .50-1.0) of the six clinicians diagnosing each case.

2. Fifty-two percent of the total cues available were collected across all four cases.

3. One percent of the same diagnostic statements in the check list were checked by three or more (P.A. = .50-1.0) of the six clinicians diagnosing each case.

4. Twenty percent of the diagnostic statements in the check list were checked for the four cases.

5. The commonality scores indicated that the clinicians for this study were consistent, 75 percent, in their agreement on cues collected across all cases.

6. The commonality scores also indicated that the subjects for this study agreed, on the average, 55 percent of the time on the diagnostic statements made across all cases and all subjects.

7. The inter/intraclinician agreement Phi indicated low agreement, with little variability between cases on cues collected and diagnostic statements made.

8. The intraclinician agreement Phi for cues and diagnostic statements was not consistent, and agreement was low.

9. The intraclinician agreement Phi on cues collected was higher than the inter/intraclinician agreement Phi on cues collected for three of the four cases.

10. The intraclinician agreement Phi on diagnostic statements made was higher than the inter/intraclinician agreement Phi on diagnostic statements made for all four cases.

11. With two exceptions, the Porter Index yielded slightly lower scores than did the Phi for both intra- and inter/intraclinician agreement on cues and diagnostic statements. The exceptions were the lower inter/intra Phi on cues for Case 2 and the lower intra Phi on cues for Case 1.

12. The clinicians for this study were lower in agreement (both Phi and Porter) on diagnostic statements made than on cues collected.

The major findings related to the <u>process</u> or the way in which the clinician diagnosed or behaved in the clinical interaction include the following.

1. The highest mean correlation for cue times occurred for Case 4, with the other three cases having somewhat lower but more consistent mean correlations.

2. The overall mean correlations for diagnostic statement time were fairly consistent, with the highest value for Case 4.

3. The average individual clinician compared with a group of clinicians diagnosing the same case showed greater magnitude in relationship for cue time than for diagnostic statement time.

4. The agreement among clinicians on the time/order in which diagnostic statements were made, as reflected by partitioned Phi coefficients, indicated very low, mostly negative, or no agreement.

5. The subjects diagnosing the four cases indicated that they used slightly more cue-directed inquiry than hypothesis-directed inquiry in data (cue) collection.

6. The subjects diagnosing the four cases indicated that they used more cue-directed inquiry than hypothesis-directed inquiry in making diagnostic statements about the cases.

7. The subjects diagnosing the four cases indicated greater use of cue-directed inquiry for diagnostic statements than for cues.

The informal product/process measures were designed to reflect, in a less formal way than in the more formal statistical measures, the second principle of the Inquiry Theory. That principle states that the behavioral interaction that occurs between a case and a clinician is determined in some probabilistic manner by the clinician's memory and strategy. To explore that principle informally, six questions based on Sherman's Model of Reading and Learning to Read (1978) were asked orally to the eight clinicians following the third or last clinic session for each clinician. The purpose of these questions was (1) to gain insight into the clinical diagnostic behavior of reading clinicians being defined in terms of (a) the data base represented in clinical memory and (b) the diagnostic routines represented in clinical strategy, and (2) to attempt to determine the source or schema that allowed the clinician's data base or cue collection to be translated into an action, e.g., the clinician's generation of hypotheses regarding the reading problem. Results of responses to the informal product/process questions seemed to indicate the following:

1. In general, the clinicians in this study did not share a clinical memory in terms of what information should be included in a "good" diagnosis.

2. In general, the clinicians in this study did not share a clinical strategy in terms of a diagnostic routine or how they went about a diagnosis.

3. The informal data generally seemed to support the statistical findings.

Discussion

Because of the exploratory nature of this study and the rather complicated interplay of the features of the Inquiry Theory with the actual diagnostic behavior of the clinicians, discussion will attempt to provide further understanding of the results obtained in this research.

Study of Product

In Chapter I, it was noted that three components comprising the Inquiry Theory directed the present investigation of clinical problem solving among reading clinicians. The three components, referred to as corollaries, are (1) the agreement of reading clinicians in collecting cues in order to reach a diagnosis, (2) the agreement of reading clinicians in making diagnostic statements, and

(3) the hypothesis-generation strategy or inquiry of reading clinicians in making a diagnosis. The first two corollaries enabled the investigation of what highly trained, experienced reading clinicians judged to be important material for diagnosing a reading problem. These results or the product of diagnosis are represented in the Inquiry Theory as clinical memory. The third corollary, dealing with hypothesis generation, provided the means for investigating the diagnostic routine or how reading clinicians go about a diagnosis. This process of diagnosis is represented in the Inquiry Theory as clinical strategy.

Previous research closely related to this study (Vinsonhaler, 1979a) indicated findings that can be further substantiated by the results obtained in the present study. When results of the Observational Studies of 1977 (OS '77) are compared with the results of this study (OS '78.3), several observations can be made regarding reading clinician agreement. (See Table 30.)

It appears that, on the average, reading clinicians for both studies showed some degree of commonality in terms of cues collected for a given case (SIMCASE). Likewise, clinicians for both studies showed, on the average, a higher agreement with themselves (intraclinician agreement) than with each other (inter/intraclinician agreement) in terms of cues collected for a given case (SIMCASE). Thus, it appears that the implications for cue agreement as postulated by the cue-agreement corollary (intra higher than inter/intra) do hold for this study as well as for the one to which it is compared.

Table 30.--Agreement statistics on cues collected and diagnostic statements made by eight clinicians in each of two studies using four simulated reading cases.

Statistic for Agreement	Cues		Diagnostic Statements	
	OS '77 Study	OS '78.3 Study	OS '77 Study	OS '78.3 Study
Commonality score				
Mean	.74	.75	.55	.55
Standard deviation	.03	.03	.03	.06
Inter/intra diagnosis correlation (Phi)				
Mean	.15	.36	-.07	.11
Standard deviation	.02	.05	.03	.04
Inter/intra diagnosis (Porter)				
Mean	.32	.34	.12	.08
Standard deviation	.03	.04	.01	.02
Intra diagnosis correlation (Phi)				
Mean	.33	.46	.13	.25
Standard deviation	.13	.19	.10	.08
Intra diagnosis (Porter)				
Mean	.41	.42	.23	.17
Standard deviation	.09	.16	.05	.05

No. of clinicians = 8 (per study)   No. of total diagnoses = 24 (per study)   SD = true standard deviation of the average.

Again, comparing the present study to the similar OS '77 study, it appears that, on the average, reading clinicians for both studies seemed to show only a limited degree of commonality (.55 and .55) in terms of diagnostic statements made on a given case (SIMCASE). Although clinicians for both studies showed, on the average, a higher agreement with themselves (intra agreement) than between individuals (inter/intra agreement) in terms of diagnostic statements made for a given case (SIMCASE), the diagnostic statement agreement was very low. It appears that the prediction that intraclinician diagnostic statement agreement should exceed inter/intraclinician agreement is supported for the two studies compared. Overall, results of the present study support the conclusion of the previous study, which is that "reading clinicians show a higher level of agreement with themselves and others in the data collected during the clinical encounter than they do in stating the diagnosis based upon such cue collection" (Vinsonhaler, 1979a, p. 25).

However, although the data for the two studies support the prediction of the agreement corollary that a clinician's agreement with himself would be equal to or greater than the agreement of individuals, the corollary does not make a prediction about the direction and the extent of any differences that might occur. An example of what could occur without having the <u>direction</u> of agreement defined would be the situation of having negative values for both intra and inter/intra agreement. For example, if the <u>intra</u> Phi mean value was -.02 and the inter/intra Phi mean value was -.002, the agreement corollary would still hold in terms of its prediction, regardless of the negative values. It would therefore seem that the corollary should be stated more clearly in terms of direction so as to account for positive, equal, and negative agreement. Additionally, although reference has been made to the possibility that low inter/intra-clinician agreement would suggest a lack of common memories (Vinsonhaler, 1979a), the Inquiry Theory may need further refinement in this regard; i.e., in terms of the Inquiry Theory, what is low agreement? Subsequently, since the corollary of agreement has implications for the Inquiry Theory in terms of the data base or memory shared by clinicians, one might raise the question: To what degree do clinicians share a memory (or strategy)? In addition, the intra-clinician agreement scores raise the question of the degree to which clinicians should agree with themselves on cues collected and diagnostic statements made. Too much consensus may reflect narrow and inflexible approaches to diagnosis, whereas too little agreement might indicate a lack of clinical expertise.

The lack of intraclinician agreement on diagnostic statements made by reading clinicians in this study could be a reflection of check-list length and complexity. It would appear that this area bears further investigation using check lists of varying lengths and formats before agreement is considered a viable measure of clinical expertise in reading diagnosis. One might wonder if the lack of agreement on diagnostic statements is a function of the specificity of the diagnostic categories used.

Whether case difficulty is a function of inter/intraclinician agreement should be considered. The commonality agreement measure for cues, for example, did not appear to be affected by the difficulty of a case since both an easy case (Case 4) and the most difficult case (Case 3) had the same score, and the scores were consistent for all four cases. The commonality agreement measure for diagnostic statements showed more variability of scores between Case 3 (most difficult) and Case 4 (easy). However, Case 1 (more difficult) and Case 2 (easiest) had almost identical scores, .55 and .54, respectively. Therefore, case difficulty may or may not have contributed to the differences in the commonality scores between cases. However, the difference between cue agreement and diagnostic statement agreement might also indicate that some factor(s) other than or in addition to case difficulty was (were) operating for diagnostic statements. One possibility is subject differences, but another, more likely variable is the length and complexity of the RDCL used to make diagnostic statements about each of the four simulated reading cases.

Both the Phi and Porter statistics as measures of inter/intraclinician agreement reflected the possibility that agreement was a function of case difficulty. The highest agreement values were at the easier end of the difficulty scale, whereas lower agreement values appeared for the more difficult cases. This was true for both cue and diagnostic statement data. It should be noted that in medical studies the commonality or group consensus had been accepted as a valid measure of agreement. However, the Inquiry Theory suggested the possibility that differences exist between individual and group agreement in diagnoses. These differences might account for the discrepancies between the commonality scores and the results based on statistics reflecting agreement between individuals (Vinsonhaler, 1979a).

Study of Process

Vinsonhaler (1979a) reported that of the two types of strategies (hypothesis directed and cue directed) currently being used to define process in the Inquiry Theory, inductive or information-directed inquiry appeared to be generally more characteristic of reading clinicians. This was in contrast to what had been observed in medicine, in which hypothesis-directed or deductive-strategy inquiry appeared to be more generally predominant. Although the methods and analyses used in this study to define process strategies were not the same as those used in the previous OS '77 study, the present study generally supports the earlier findings and may help in making the nature of the phenomenon more clearly understood.

One of the decisions that had to be made in this study (OS '78.3) was on what basis a subject, in terms of his responses to selected questions, could be characterized as having used hypothesis-directed or cue-directed inquiry; this problem may not yet be resolved. Certainly, the degree of difference needed for the strategy to be characterized as hypothesis directed or data directed has not been defined. At present, however, one can look at the averages in Tables 21 and 24 and make a tentative, partial judgment about hypothesis-directed inquiry by noting the consistency of the scores for the three cases. Subject 101, for example, was fairly consistent for both cues and diagnostic statements. Contrary to what might have been expected, Subject 105 also had consistent mean values for average of percentage of number of times relationships appeared for cues and statements. Subject 105 had been consistently low in terms of self-agreement, with an intraclinician Phi agreement of .38 for cues and .01 for diagnostic statements. This would imply that a careful look at subject differences might reveal helpful information about individual reading clinicians' patterns of clinical diagnostic behavior as well as provide a better understanding of the measures used to describe behavior relative to the Inquiry Theory.

Biasing Effects

In making comparisons between the results of the OS '77 study and the OS '78.3 study, a number of factors that may account for a portion of the differences between the two studies as well as point out possible biasing effects in this study should be mentioned. Those factors involve (1) procedures of data collection; (2) instrumentation, e.g., RDCL; and (3) size of the cue and diagnostic statement domains.

Procedures.--The decision not to use a data or cue inventory for each case in this study was made to determine more reliabily the extent of clinical memory for the eight subjects in the study. The researcher felt the use of a cue list might cause subjects to request data they might not usually collect just because it was available, and that this would thereby artificially inflate the clinician agreement for the cues collected. It was also believed that lack of a cue inventory would result in a more realistic indication of what tests the clinicians might administer or the information they might gather if they were working with a child rather than a SIMCASE. In other words, it seemed desirable not to introduce the "power of suggestion" by using a data inventory.

However, it should be noted that not having the inventory presented several problems. For example, some of the information requested was not available for a particular case, though it might have been available for another case; this, of course, meant a loss of time as the clinicians attempted to request specific information that was available. However, if the clinician asked for a test that was not available, he could then request a general type of information and would be told what similar test was available. Every attempt was made not to play a "guessing game" with the subjects. The experimenter had a data inventory to facilitate easy location of material in the SIMCASE box.

Although interesting information might have been provided on how the clinicians in this study used their clinical memories to collect data for simulated reading cases, a biasing effect was created by not providing the subjects with inventories. The use of data inventories would have tightened controls and further standardized procedures, especially since the task of cue collection was timed. One must weigh the tradeoffs because results of agreement for two studies, one using a data inventory and one not, are similar. (See Table 30.) Without the use of decision aids, i.e., a data inventory, the clinicians in the OS '78.3 study did as well as those in the earlier study and in general improved in their agreement on the cues requested. It should be noted, however, that agreement in OS '78.3 was computed on the cues or data available for each case regardless of the cues requested. A list of the cues requested for each case but not available appears in Appendix D. A comparison of procedures for the two studies (OS '77 and OS '78.3) is presented in Appendix G.

Instrumentation.--Although it may be noted that there was some increase of clinician agreement in the study reported in this dissertation over clinician agreement in the OS '77 study, the increase was not as great as might have been expected when the study began. It was felt that introducing the standardized RDCL to be completed by the subjects as a more objective measure would eliminate those biasing effects that might have been created by having someone other than the subject himself interpret the written diagnosis. What appears to have happened is that the length and complexity of the RDCL used in this study affected to some extent the clinician agreement on diagnostic statements. One of the reasons for the length and complexity of the list was to make it as comprehensive as the diagnosis itself should be. Even so, the list does not reflect relationships between variables affecting reading, e.g., "Reliance on context helps his comprehension." The effects of check-list length and complexity should be the subject of future investigation.

In regard to the check list itself, the subjects' comments that were listed in "Other" on the check list are presented in Appendix D. Some of these comments should have been checked on the list because they had corresponding items on the check list. These items would then have been included in the agreement analyses. Other comments should not have been considered diagnostic statements according to the definition provided in the instructions to the subjects and should not have been written in. Still other comments were repetitions of diagnostic comments that had already been checked in the RDCL. It should be recalled that subjects were asked to avoid using "Other" categories if possible. An instruction to subjects to go back and try to fit "Others" comments into the check list might have resulted in a few, though likely insignificant, changes in clinician agreement. Because there is no standard terminology in reading diagnosis, clinicians might not have been able to locate items in the check list that were comparable to their own diagnostic statements. This problem with terminology might also have prevented a match in clinician agreement.

The measures of hypothesis-directed inquiry as opposed to information or cue-directed inquiry used in this study were pilot attempts to understand the intellectual process better. However, one might speculate about the implications in general terms of the higher (above 50) or lower (below 50) scores of the clinicians on a scale from 0 to 100. For example, a score of 50 or above would indicate that the clinicians had used 50 percent or more of either information (cue)-directed or hypothesis-directed inquiry. Factors unique to the data collection and subsequently the input to the computer resulted in some problems that need to be resolved before data analysis using the same analysis program for cue-to-statement relationships in the future.

Size of domains.--The analysis of the data for cues in both studies (OS '77 and OS '78.3) required a cue domain or list of information for each case. Since each of the four SIMCASEs contained different data, each case had its own cue domain. Although the total cue domains for each case were the same in both studies, the analyses were calculated differently. In OS '77 the analysis for cues was computed only on those cues that were requested at least once, whereas in the OS '78.3 study the analysis for cues was computed on the total number of cues available for each SIMCASE. Table H24 in Appendix H shows the mean and standard deviation of the Porter statistic, including the number of cues and diagnostic statements in the domains, for the OS '77 and OS '78.3 studies.

The diagnostic statement domain or RDCL for OS '78.3 included 507 items, and the agreement was calculated for all four cases using the same domain. The OS '77 study data were calculated on the number of diagnostic statements mentioned for each case by at least one of the six clinicians diagnosing that particular case.

The effects of domain size or using the total domain as opposed to only those items mentioned to compute analysis should be studied.

Conclusions

Regardless of the small sample of reading clinicians participating in this study, there is evidence to support a number of conclusions. These conclusions are:

1. Experienced reading clinicians using simulated reading cases appear not to share a common data base (memory) regarding what information (cues) should be included in a diagnosis or what diagnostic statements are important in writing a diagnosis.

2. Experienced reading clinicians using simulated reading cases appear not to share a common diagnostic routine (strategy) in terms of how to go about a diagnosis.

3. Experienced reading clinicians using simulated reading cases appear not to reflect the consistent use of a theoretic process model of reading diagnosis as might be reflected in hypothesis-directed inquiry.

These conclusions will be discussed in terms of (1) an overriding or general factor related to reading diagnosis, that of training; and (2) factors more specifically associated with this particular research effort, the RDCL and the case complexity.

One major factor related to reading diagnosis appeared to be suggested in the literature and was manifested by both the formal and informal measures of agreement used in this study. That factor was training.

Training

The influence of training or, more accurately, the lack of formal training for clinicians in reading diagnosis, is evident from the literature. Generally mentioned in discussions of diagnostic training in reading are various sets of principles of diagnosis, many of them sequential but few of them in agreement, either in terms of what information should be included in a diagnosis or how a diagnosis should be conducted. Missing is a set of standard clinical reading cases with which students may learn and practice their clinical skills. This probably means that the question of case complexity relative to information requests and hypothesis testing has not routinely been dealt with in diagnostic reading instruction courses. Included in the literature are discussions dealing with cause/effect relationships in reading diagnosis, but primarily in a nonspecific sense; i.e, "Rick's poor sight vocabulary is due to his frequent absences from school in first grade." Lacking are discussions of treatable cause/effect relationships; i.e., "Mary's failure to analyze and pronounce unknown words indicates that she is probably relying on sight-word recognition and she needs instruction in using phonetic and contextual analysis."

The apparent lack of training in systematic, thorough, and comprehensive clinical diagnosis in reading might be reflected in the generally low diagnostic agreement of clinicians in this study. Perhaps more revealing, however, is the fact that clinicians confined their diagnostic statements to approximately 17 to 22 percent of the check list across all cases. This would seem to reflect the lack of comprehensive diagnostic strategy. Not to be ignored are the results of the informal measures, which further support the previous conclusions. The clinicians in this study did not agree on the content of a "good" diagnosis, nor did they agree on the way in which a diagnosis should be conducted. Likewise, they did not indicate the consistent use of a theoretical process model of diagnosis.

Two additional factors related more specifically to this research and more convincingly supported by both formal and informal results are (1) the factors influencing the clinicians' use of the RDCL and (2) the elements of diagnosis affected by problem or case complexity.

Check List

Three things that possibly influenced the clinicians' use of the RDCL were (1) the length and complexity of the check list, (2) the specificity of the diagnostic categories, and (3) the lack of standard diagnostic terminology that could be used in the check list.

The effect of check-list length and complexity was evidenced by the low reliability of clinician use of the RDCL and by the much lower agreement of the clinicians on diagnostic statements made than on cues collected for specific reading cases. Another influence may have been the specific nature of the diagnostic items and the necessary overlap of these items within the major categories. For example, "Vision--General Statement" appeared under the category Developmental/ Physical Factors, whereas "Visual Discrimination--General" appeared under the category Perception/Perceptual Motor/Language Factors.

The literature is replete with examples of disagreement on a standard terminology for reading. This conflict may have extended to this study through the clinicians' interpretations of items in the check list. Additionally, although the reading literature sometimes indicates the desirability of having clinicians determine the strengths and weaknesses of students to guide in effective remediation, there is no clear evidence that training programs teach this diagnostic skill. The present study indicated that clinicians were often confused about what constituted a strength, a weakness, or just an observation for a student, in spite of the fact that these terms were carefully defined. Some clinicians, for example, checked both strength and weakness for the same diagnostic item. It appeared that the formal identification of strengths and weaknesses as part of reading diagnosis may have been a diagnostic skill with which they were not trained or practiced.

Recommendations

As in any new and developing area of research, conclusions must remain tentative lest evidence be used without proper support. In medicine, the example that comes to mind is the use of new drugs that have not been thoroughly tested. In education, the concern is more likely to be that even with what would seem to be supporting evidence, it frequently takes years for changes to be implemented. Researchers working with the Inquiry Theory in terms of testing its corollaries or working on the theory in terms of its explication are concerned about both implications: that too little evidence may produce false positives (or false negatives) and that accumulating too much evidence may complicate the theory to the point that educators will see little practical value for its use. It is this challenge of circumstances that has provided one impetus for both the care and speed with which Inquiry Theory researchers have almost doggedly pursued their common interest.

Many recommendations come to mind as a result of one's having worked to explicate a new theory. Several of these suggestions follow.

Like any new theory in which attempts are made to verify or establish an explanation of phenomena, the Inquiry Theory needs refinement. One concern that becomes of paramount importance when attempting to account for behavioral phenomena is the same concern that was expressed earlier in this writing with regard to reading, that of confusion over terminology. In the explanation of the Inquiry Theory, four terms appear presently to be clouding succinct exposition. These terms are "corollaries," "hypotheses," "principles," and "assumptions." Each needs to be more clearly defined in terms of its relation to the others and to the theory.

A second recommendation regarding the theory is somewhat related to the first: Whereas "memory" has been described or defined in the theory in terms of what it might comprise (problem, cue values, prescriptions, and treatment descriptions, and the relations between them), it has not been clearly defined in behavioral terms. Can it be said, for example, that the data base (memory) of a reading clinician includes experiences, and if so would this data base not therefore need to be flexible, depending on the particular problem being diagnosed? Do reading clinicians hold a particular causal theory on the basis of empirical observation of covariation between stimuli and responses? Is it possible for a reading clinician to be unable to detect a relationship or covariation because he lacks a hypothesis (theory) leading him to suspect covariation? Conversely, is it possible for a reading clinician to perceive covariation where none exists because he has a hypothesis (theory) that leads him to expect it? It would appear that to the extent diagnosticians share similar connotative networks, they would be expected to arrive at similar judgments about the likelihood of a causal link between stimulus and response. An example of a connotative or secondary associated meaning in a reading diagnosis might be: "He doesn't seem to know many sight words and he confuses many words that are similar. I'll check his visual memory and visual discrimination."

Another problem that needs to be addressed must necessarily involve more clinicians with similar training and background diagnosing several cases representing different reading problems. The question one might then answer would be: Is diagnostic success on some reading problems correlated with success on other problems; i.e., is diagnostic accuracy case specific? The results of this study indicate that diagnostic strategy might be affected by case difficulty.

It might be recalled that in the clinical session for this study the clinicians were asked to write both a diagnosis <u>and</u> a remediation. Because of the massiveness of the data and their analysis, it was decided to limit the examination to only that portion concerning diagnosis, although for this researcher diagnosis and remediation are inseparable. Subsequently, it would seem appropriate to look at the remediation data from this study in light of diagnostic results.

Along the lines of remediation, one question that might be relevant asks: Is it possible to store (in long-term memory) all of the alternative remediations available for a given diagnostic problem? In answering this question, multi-causal diagnosis and therefore multi-remediation must be considered. Following from this, one might suggest the questions: Can reading clinicians "nest" problems such as physicians do for infections? Do reading clinicians have a treatment memory and strategy?

Of additional research import would be:

1. the identification of specific distinctive features of individual reading cases,

2. the study of acquisition and interpretation of cues relevant to the correct and/or highest priority diagnostic hypotheses as well as the types of diagnostic hypotheses generated,

3. the study of how the patterns of hypothesis generation vary from case to case or clinician to clinician,

4. the study of the effect of knowledge and experience in solving specific reading problems in clinical diagnosis,

5. the study of clinical decisions made regarding relevant as opposed to irrelevant sources of data or information used for diagnosis of reading problems, and

6. the development of precisely documented descriptive studies of diagnostic strategy in reading diagnosis using real cases for the purpose of studying diagnostic intellectual process.

In summary, there are several converging lines of research, each having promise for the study of clinical problem solving as it applies to reading clinicians. This study, although limited in its implications, was one attempt to diminish the gaps between theoretic and methodologic development of the Inquiry Theory of Clinical Problem Solving as it relates to clinical reading diagnosis.

APPENDICES

APPENDIX A

CLINICIAN BACKGROUND FORM AND INFORMATION

BACKGROUND INFORMATION ON READING CLINICIANS

Name: ______________________________

Home Address: ______________________________

____

Office or School Address: ______________________________

____

Phone--Home: ______________ Office or School: ______________

Academic Background:

Institutions Attended	Major	Degrees Received	Dates

Additional Coursework: ______________________________

____

____

Teaching Experience--Grade Level: ________ No. Years: ________

________ ________

________ ________

Total Years' Diagnosing/Consulting Experience: ______________

Present Position:

Title: ______________________________

Location: ______________________________

Responsibilities: ______________________________

____

____

Years in This Position: ______________

Other Relevant Educational Experience: ______________________________

____

____

____

Table A1.--Background information on reading clinician subjects.

Subject Number	Degrees Obtained	Major Area of Study	# of Years Teaching Experience	Grade Level of Teaching	# of Years Diagnosis Experience
101	B.A. M.A. Ph.D.	Elem. Ed. Reading Reading	10	Grades 2, 4, & 5 College	2 part time 2 full time
102	B.A. M.A. 80% on Ph.D.	Elem. Ed./Reading Reading Reading	7	Grades 3, 4, & 5 College	4 part time
103	B.A. M.A. Ph.D.	Elem. Ed. Elem. Ed./Reading Elem. Ed./Reading	5	Grades 3, 4, & 5 College	4 full time
104	B.A. M.A. 66% on Ph.D.	Poli. Sci. Elem. Ed. Reading	7	Grades 1, 4,5,6,7,&8 College	3 full time

Table A1.--Continued.

Subject Number	Degrees Obtained	Major Area of Study	# of Years Teaching Experience	Grade Level of Teaching	# of Years Diagnosis Experience
105	B.A. M.A.+	English Education	5	Grades K & 9-12 College	15 full time
106	B.A. M.A. Ed. Specialist (Also Spec. Ed. & L.D. approval)	Elem. Ed. Reading Reading	5	Grades 1-3 College	12 full time
107	B.A. M.A. (Also working on Ph.D.)	Elem. Ed. Ed. Psy. Reading	4	Grades 1&2 College	11 full time
108	B.A. M.A. (Also additional work on Ph.D.)	Psy./Ed. Reading Learn. Disabl. & Administration	7	Grades 4, 6, 7, & 8 College	13 full time

APPENDIX B

CLINICAL SETTING AND INSTRUCTIONS

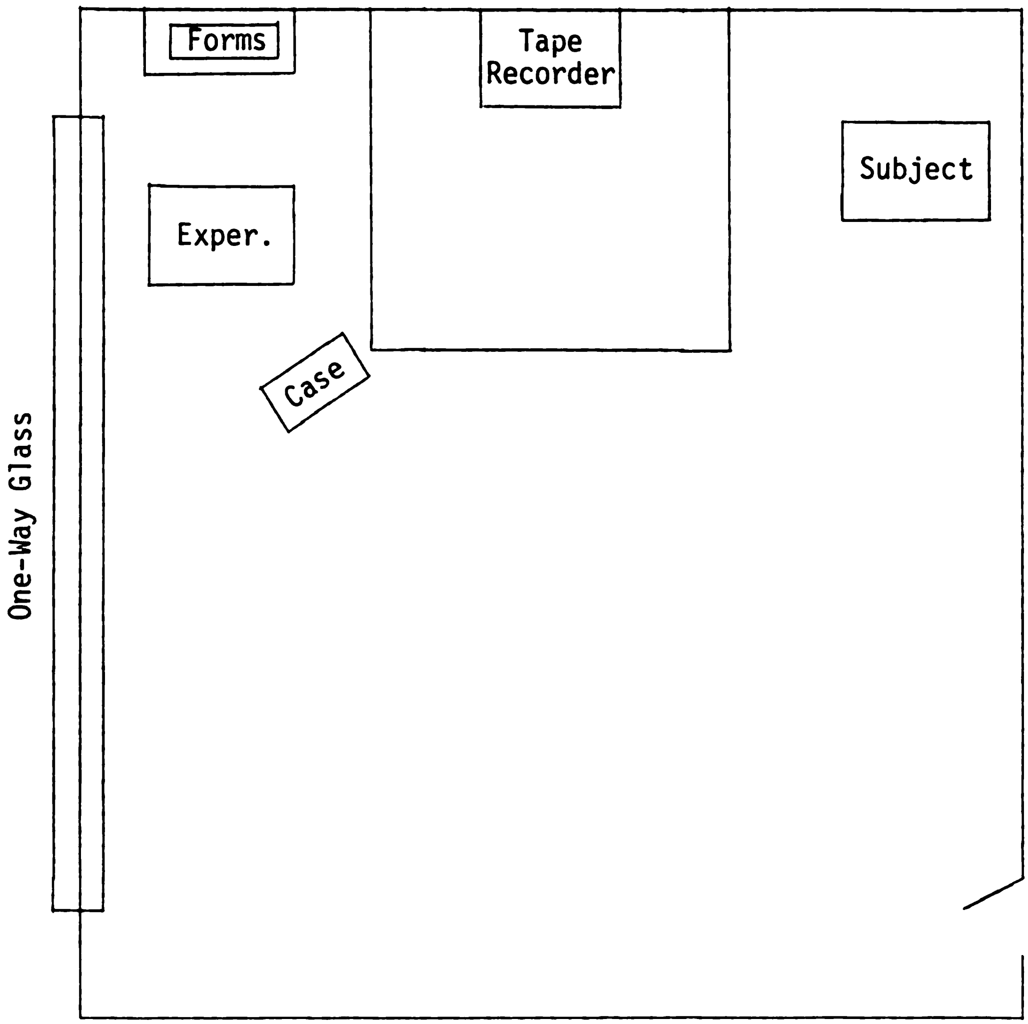

Figure B1.--Diagram of clinical setting.

INSTRUCTIONS TO SUBJECTS FOR CLINICAL SESSION

This research is being carried out by the Institute for Research on Teaching at Michigan State University. The Institute needs to observe a number of representative reading clinicians in order to develop theories and computer simulations of how clinicians diagnose reading problems. You have been chosen as one of those who will be observed during our study. Because this work will take several hours, the IRT will pay you as a consultant for the College of Education. A check will be mailed to you after the session.

Before I explain the observational session, I should emphasize that all personal information regarding this session will be kept confidential. Your name will not be part of Institute permanent records. Instead, a number will be used. We are required by law to protect your privacy by keeping confidential your name, social security number, etc. Second, I should emphasize that we are not evaluating you in any way. We are merely interested in understanding how you usually go about determining the most probable reading problems of a given client.

Now I will explain what we will be doing. The session will be divided into three parts. In Part I you will be asked to interact with materials to analyze a case of reading difficulty. The case materials with which you will be working will be in this file box (indicate box). The case information can be provided in these forms (show list of <u>Forms of Information</u> to subject): (1) test scores (show example), (2) test booklets (show example), and (3) examiner's comments (show example). (4) There are also audio recordings of reading test sessions (show example).

For Part II, you will be asked to write a diagnosis for this case and then to write suggestions for appropriate treatment or remediation based on your diagnostic findings.

In the final part of the session, Part III, you will be asked to transfer your written diagnosis to a Diagnostic Check List. Also during Part III, you will be asked to attempt to recall what you were thinking about as you worked on the diagnosis. Here is an overview of what we will be doing for the three parts of this session. (Indicate Session Overview.) I will explain each part in more detail as we come to it.

We will begin now with Part I. Your task is to request the information about a case which will be used to determine the most likely diagnosis and to suggest a general program of remediation. There is no right or wrong amount of information to request for your diagnosis. I would like you to diagnose this case in much the same manner you would use in diagnosing a real case. Assume that you are working with the child in a one-to-one setting. You will be given 45 minutes in which to request information on the case. When you request an item of information, I will give it to you. You may keep all items of information throughout the session. Request items in the order in which you normally collect such information. You may take notes if you wish. (Indicate note pad.) I will begin timing when you make your first request for information and I will let you know when there are 20 minutes of the time remaining.

To review, here is a summary of the instructions for Part I. (Give subject Summary of Part I Instructions.):

1. As a consultant, you have been called in to examine and analyze a reading case.

2. Ask for needed information as you would normally collect data.

3. You may take notes if you wish.

4. You will have 45 minutes to reach a decision on the diagnosis.

5. If you reach your diagnosis before the time is called, indicate that to me.

Are there any questions?

I will have a tape recorder turned on during the session just in case there is something to which I might need to refer at a later time.

Now here are some initial items of information on the case before we begin timing. (Give picture and initial contact material. Start tape of initial contact. Make sure tape is at starting point.)

You may begin your request for further information on the case when you are ready.

What information would you like first?

(Begin timing before and after IC and when subject makes first request for information.)

(Record time of cue request under column heading labeled CUE REQUEST TIME on large blue sheet--Hypothesis/Observation Check List.)

(Write cue name of cue requested under column heading labeled CUE REQUEST ORDER on large blue sheet--Hypothesis/Observation Check List.)

(Twenty-five minutes after first cue request, remind subject he has 20 minutes of time remaining--say, "You have 20 minutes left.")

(After subject has completed his work or time is called, give him NCR Blank Form paper, two sheets, and instruct him as follows.)

Now for Part II, I would like you to summarize your judgments in written form. Please briefly state your diagnostic opinions on this special carbon paper. (Indicate double sheet of paper on clipboard--DIAGNOSIS.) Write only on the top sheet. Additional paper is available should you need it. In writing your diagnosis, please write as clearly as possible and double space between each line. Be as specific as you can and use complete sentences. In writing your diagnosis, assume that the report will be used by a clinician with training similar to yours who will work one-to-one with the student. You will have 25 minutes to write your diagnosis. Here is a copy of the instructions. (Give subject copy of the <u>Instructions for Writing the Diagnosis</u>. Allow time for him to read instructions.) Are there any questions? You may begin.

(Begin timing.)

(When subject has finished or time is called, place his name, the date, and the SIMCASE name on the Diagnosis write-up.)

Now please write your suggestions for remediation here on the special carbon paper. (Indicate Remediation clipboard.) You may use whatever form is convenient for you in writing your remediation. You will have another 25 minutes in which to complete this task. Are there any questions?

(When the subject has finished writing the remediation or time is called, place his name, the date, and the SIMCASE name on the remediation write-up.)

(Give the subject a 10-minute break.)

(During the break the Examiner should put all cues in the order in which they were requested by the subject. Separate the double sheets of carbon paper for the diagnosis write-up and number the pages in correct order. File one complete copy of the diagnosis in the subject's folder and leave the other copy on the table.)

We will now begin Part III and the final part of this session.

In order to help me objectify data and to make sure that I understand your diagnosis, I would like for you to transfer your written diagnosis to a check list. This list is made up of possible diagnostic statements which might apply to students with reading problems. The statements in the check list are restatements of actual statements made by reading diagnosticians that have been put into a standard vocabulary. The statements are divided into several categories such as Developmental/Physical Factors, Perception/Perceptual Motor/Language Factors, Reading-Isolated Instant Word Recognition Factors, etc. These categories listed on a cover sheet and having section tabs should make it easier for you to locate statements for transferring your diagnosis. There are more statements on the check list than you will need and you might not have statements in all of the categories. Please take a few minutes now to look over the check list and then I will give you further instructions. (Give subject sample check list. Allow time for the subject to look over the check list.)

Now, to make sure that you understand what you are to do, I will first show you an example of how to transfer a written diagnosis to a sample portion of a check list. Then I will give you a chance to practice a transfer using another example.

This is a sample portion of an actual written diagnosis for a reading case. (Show Sample #1--Diagnosis.) It is not meant to be a model or to indicate how a diagnosis should be written. Notice now that circles have been drawn around diagnostic statements. A diagnostic statement is the identification of a factor or variable which helps determine the state or condition of a student's reading performance. It may be one sentence or several and it may also just be part of a sentence. In addition, the diagnostic statement may be a strength, a weakness, or just an observation. The circled diagnostic statements or reading factors are numbered, as you can see in the sample. Next you can see how the circled and numbered factors have been transferred to a sample part of the diagnostic check list. (Indicate Check List Sample #1.) The number of the diagnostic statement in the written diagnosis is placed to the left of the matching diagnostic statement on the check list and under the appropriate column to indicate that the statement or factor represents either a strength, a weakness, or is just an observation. (Indicate number on Sample Check List #1.) Are there any questions? (Collect sample materials.)

Here now is another sample from a written diagnosis. (Indicate Practice Sample #1--Diagnosis.) Please locate, circle, and number one diagnostic or reading factor statement in this sample. (Allow time for task.) Here is a copy of the Diagnostic Check List. (Present practice copy of Diagnostic Check List.) Now, using the check list, locate the main category into which you believe the circled statement from the diagnosis will best fit. Use the cover sheet of the check list to help you decide upon and locate the appropriate main category of the check list. (Allow time for task.) Next locate under the main category heading which you have chosen the statement which most nearly matches the statement from the written diagnosis. (Allow time for task.) Put the number of the circled diagnostic statements to the left of the corresponding statement on the check list and under the appropriate column heading--strength, weakness, or observation. (Allow time for task.) Note that space is provided within each main category for "Other Statements." You may copy any statements from the written diagnosis under "Other Statements" if you are able to determine the main category but are unable to find a matching statement on the check list in that same category. However, statements in the written diagnosis for which you can not determine a category should be copied under the main category Roman Numeral #10 OTHER FACTORS in the check list. (Indicate X. OTHER FACTORS.) Please make as limited use as possible of the statements and the category called Other. Are there any questions? (Collect practice materials.)

Here is a review of the instructions. (Give copy to subject.)

1. Locate all of the diagnostic or reading factor statements in your written diagnosis.

2. Circle and number the diagnostic statements.

3. Locate on the cover sheet of the Diagnostic Check List the main category into which your diagnostic statement would most likely fit.

4. Find that main category in the check list and locate a statement within it which best matches your circled statement.

5. Decide if your statement indicates a strength, a weakness, or is just an observation which you made; then place the number of your circled statement under one of the column headings.

6. If you recall information that you did not include in your written diagnosis, write that information on the Stimulated Recall sheet, then transfer it to the check list using A, B, C, D, etc.

In transferring your diagnosis, it is important that you do not add diagnostic statements from memory to the check list even if something should come to mind. You may, however, write those thoughts on this sheet of paper. (Indicate Stimulated Recall tablet.) Here is a copy of the Diagnostic Check List and your written diagnosis. I would like for you to transfer your written diagnosis to the check list according to the instructions we just reviewed. You will not be timed on this task, but work as rapidly as possible. You may begin.

(If subject adds diagnostic statements to the Stimulated Recall sheet, remind him that those statements are to be added to the check list by using A, B, C, D, etc. in the columns.)

To complete our work, I would like to have you clarify for me the way you went about making your decisions for the diagnosis. To help you with recall, I have recorded the information you requested in the order in which you requested it here on this sheet. (Indicate large blue sheet--Hypothesis Observation Check List.) I will tell you the name of the information and then I would like for you to complete two main statements by checking and then explaining your responses.

Here is a sample (indicate Sample #2--Hypothesis/Observation) of the Hypothesis/Observation Check List. It is not intended to represent a model. In other words, the statements in the sample under the columns EXPLAIN are not to suggest the way your statements should be worded. Notice that in some instances more than one answer has been checked. (Indicate fifth column for third example.) You may check more than one answer whenever you feel it is appropriate. Be sure to explain each time you check a response. Do you have any questions about the sample?

Now I will give you a chance to practice using the Hypothesis/Observation Check List form. (Give subject practice Hypothesis/Observation Check List.) Three items of possible information on a reading case have been filled in under column 2. Assume that you requested one of the pieces of information listed and complete the practice for a hypothetical case using that one piece of information. If you have questions, be sure to ask them. (Allow time for task.)

Now I will give you your list of requested information. (Give subject his own Hypothesis/Observation Check List.) I will show you the information in the order in which you requested it. You are to check your response in column 3, then explain it in column 4. Then check your response in column 5 and explain it in column 6. Use the same procedures for each piece of information you requested when diagnosing the case. You will not be timed on this task. (Allow time for task.) That completes this session.

(See Unedited Dialogue in Appendix D for example of additional format followed in each third or last session.)

APPENDIX C

CLINICAL OBSERVATION INSTRUMENTATION

READING DIAGNOSTIC CHECK LIST*

DEVELOPMENTAL/PHYSICAL FACTORS	I.
SOCIAL/PSYCHOLOGICAL FACTORS	II.
EDUCATIONAL FACTORS	III.
PERCEPTION/PERCEPTUAL MOTOR/LANGUAGE FACTORS	VI.
READING-CONTEXTUAL FACTORS	V.
READING-ISOLATED INSTANT WORD RECOGNITION FACTORS	VI.
READING-WORD ANALYSIS IN ISOLATION FACTORS	VII.
READING-ORAL/SILENT FACTORS	VIII.
READING-COMPREHENSION FACTORS	IX.
OTHER FACTORS	X.

*Developed by Linda Patriarca, Joel VanRoekel, George Sherman, and Ethelyn Hoffmeyer.

Strength	Weakness	Observation	
			I. DEVELOPMENTAL/PHYSICAL FACTORS
			General Health
			General Physical Development
			Physical Coordination
			Physical Activity Level
			Physical Activity-Sports
			Vision-General Statement
			Vision-Acuity
			Hearing-General Statement
			Hearing-Acuity
			Allergies
			Birth Process
			Neurologic
			Speech Production
			Other Statements:
			II. SOCIAL/PSYCHOLOGICAL FACTORS
			Intellectual/Educational Potential-General/Overall
			Verbal Intellectual Potential
			Nonverbal Intellectual Potential
			Verbal Performance Compared to Nonverbal Performance
			Potential for Grade Level Work-General Statement
			Potential for Grade Level Work-Reading

Strength	Weakness	Observation	
			Home Background-General Statement
			Home Background-Sibling Relationships
			Home Environment-Influences on Academics
			Home Environment-Influences on Reading
			Attitude Toward School-General
			Attitude Toward Reading-Instructional
			Attitude Toward Reading-Independent
			Classroom Behavior
			Relationship With Peers
			Cooperation in Group Activities
			Ability to Work Independently
			Ability to Work in a One-to-One Situation
			Social Adjustment-General
			Level of Responsibility
			Aggressiveness
			Emotional Adjustment-General
			Confidence in Own Ability-Academic
			Maturation
			Variety of Interests
			Willingness to Participate in Competitive Activities
			Ability to Deal With New Situations
			Ability to Retain Information
			Attending Behavior-General Statement
			Attending Behavior-Getting to Attention
			Attending Behavior-Selecting and Organizing

Strength	Weakness	Observation	
			Attending Behavior-Pausing and Reflecting
			Attending Behavior-Maintaining and Sustaining
			Appropriateness of Verbalizations
			Socio-Economic Status
			English as a Second Language
			Other Statements:
			III. EDUCATIONAL FACTORS
			Grade Level Placement-General Statement
			Motivation-For Academic Work
			Motivation-Reading
			Progress in School-General Statement
			Progress in School-Reading
			Quality of Instruction
			Instructional Materials-Appropriateness-General Statement
			Instructional Materials-Appropriateness-Reading
			Completion of Assignments
			Rate of Work
			Accuracy of Work
			Amount of Practice Required in Subject Matter Areas
			Level of General Information
			Other Statements:

Strength	Weakness	Observation	IV. PERCEPTION/PERCEPTUAL MOTOR/ LANGUAGE FACTORS
			Perception-General
			Perception-As Related to Academic Growth
			Auditory Memory-General
			Auditory-Sounds
			Auditory Memory-Words
			Auditory Memory-Sentences
			Auditory Sequencing
			Auditory Discrimination-General
			Auditory Discrimination-Sounds in Words
			Auditory Discrimination-Whole Words
			Visual Perception-General
			Visual Memory-General
			Visual Memory of Words-Recognition
			Visual Sequential Memory
			Visual Sequencing
			Visual Discrimination-General
			Visual Discrimination-Shapes
			Visual Discrimination-Letters
			Visual Discrimination-Whole Words
			Visual Association-General
			Visual Skills in Relation to Auditory
			Motor Development-General
			Motor Development-Onset of Walking
			Gross Motor Coordination Skills

Strength	Weakness	Observation	
			Fine Motor Coordination Skills
			Ability on Paper-Pencil Tasks
			Language-General
			Verbal Skills (Syntax)
			Articulation
			Vocabulary-Oral
			Vocabulary-Reading
			Listening-Receptive Language-General
			Listening-Ability to Comprehend Spoken Language at Grade Level
			Other Statements:
			V. READING-CONTEXTUAL FACTORS
			Oral Reading-General
			Oral Reading-Score
			Oral Reading-Accuracy/General
			Oral Reading-Hesitations
			Oral Reading-Insertions Contextually Acceptable
			Oral Reading-Miscues Contextually Acceptable
			Oral Reading-Omissions Contextually Acceptable
			Oral Reading-Punctuation
			Self-Correction of Oral Reading Errors
			Silent Reading-General
			Silent Reading-Score

Strength	Weakness	Observation	
			Word Recognition-Contextual
			Word Analysis-Contextual
			Independent Reading Level
			Instructional Reading Level
			Frustration Reading Level
			Frequency of Independent Reading
			Application of Isolated Skills to Contextual Reading
			Reading Performance Relative to Grade Placement
			Other Statements:
			VI. READING-ISOLATED INSTANT WORD RECOGNITION FACTORS
			Word Recognition-General
			Word Recognition-Basic Sight Words
			Basic Sight Word Score
			Utilization of Whole Word Approach
			Ability to Deal With Irregular Spelling Patterns
			Consistency of Error Patterns in Word Identification
			Other Statements:
			VII. READING-WORD ANALYSIS IN ISOLATION FACTORS
			Word Analysis-General

Strength	Weakness	Observation	
			Phonetic Analysis-General
			Use of Initial Consonant Sounds
			Use of Final Consonant Sounds
			Use of Blends-General
			Use of Blends-Specific
			Use of Digraphs-General
			Use of Digraphs-Specific
			Use of Vowels-General
			Use of Vowels-Specific
			Use of Vowel Pattern
			Use of Vowel Variant Pattern
			Structural Analysis-General
			Use of Prefixes
			Use of Suffixes
			Use of Word Chunks or Roots
			Use of Word Families
			Use of Phonograms
			Use of Syllables
			Ability to Decode Polysyllabic Words
			Ability to Blend Component Parts
			Ability to Blend Component Parts Auditorally
			Integration of Analysis Skills
			Ability to Pronounce Nonsense Words
			Other Statements:

Strength	Weakness	Observation	
			VIII. READING-ORAL/SILENT FACTORS
			Oral Reading-General
			Silent Reading-General
			Fluency in Oral Reading
			Fluency in Silent Reading
			Rate of Reading-General
			Rate of Reading-Oral
			Rate of Reading-Silent
			Rate of Oral Reading With Respect to Accuracy
			Use of Context-General
			Use of Context to Determine Word Pronunciation
			Use of Context to Determine Word Meaning
			Influence of Sight Vocabulary on Reading Rate
			Influence of Decoding Ability on Reading Rate
			Appropriateness of Intonations (phrases & clauses) in Oral Reading
			Other Statements:
			IX. READING-COMPREHENSION FACTORS
			Comprehension-General
			Comprehension of Grade Level Materials-Oral
			Comprehension of Grade Level Materials-Silent
			Influence of Decoding on Comprehension
			Influence of Knowledge of Word Meanings on Comprehension

Strength	Weakness	Observation	
			Influence of Interest on Comprehension
			Comparison of Oral to Silent Reading Performance-General
			Comparison of Oral to Silent Reading Performance-Score
			Recall of Sequential Information
			Use of Specific Strategies for Comprehension
			Other Statements:
			X. OTHER FACTORS

HYPOTHESIS/OBSERVATION CHECK LIST

Diagnostician_______________

Case___________ Date_________

(1) Cue Request Time	(2) Cue Request Order	(3) Did you ask for this information because---	(4) Explain.	(5) Did this information---	(6) Explain.
		--- you had a hunch? --- you just wanted this information? --- you usually get this information? --- Other?		--- confirm a hunch? --- disconfirm a hunch? --- suggest a hunch? --- Other?	
		--- you had a hunch? --- you just wanted this information? --- you usually get this information? --- Other?		--- confirm a hunch? --- disconfirm a hunch? --- suggest a hunch? --- Other?	
		--- you had a hunch? --- you just wanted this information? --- you usually get this information? --- Other?		--- confirm a hunch? --- disconfirm a hunch? --- suggest a hunch? --- Other?	
		--- you had a hunch? --- you just wanted this information? --- you usually get this information? --- Other?		--- confirm a hunch? --- disconfirm a hunch? --- suggest a hunch? --- Other?	

FORM FOR QUESTIONS ASKED FOLLOWING LAST CASE IN INFORMAL DEBRIEFING

Before we end the session, I have a few questions I would like to ask you about your opinion regarding some aspects of reading diagnosis.

1. In your opinion what kinds of information should be included in a "good" diagnosis?

2. How do you usually go about a diagnosis?

Thinking in terms of the case with which you have just worked--

3a. Do you think the student's instant word recognition was low?

Yes No

b. How do you know?

___ Reasoned in terms of relationships (cause/effect--pupils who have this problem do this [behavior]).

___ Reasoned in terms of experience with other students.

___ Reasoned in terms of authority (general reference to authority book, course, etc.).

___ Other.

4a. Do you think the student's word analysis skills were low?

Yes No

b. How do you know?

___ Reasoned in terms of relationships (cause/effect--pupils who have this problem do this [behavior]).

___ Reasoned in terms of experience with other students.

___ Reasoned in terms of authority (general reference to authority book, course, etc.).

___ Other.

5a. Do you think the student's reading comprehension was low?

Yes No

b. How do you know?

___ Reasoned in terms of relationships (cause/effect--pupils who have this problem do this [behavior]).

___ Reasoned in terms of experience with other students.

___ Reasoned in terms of authority (general reference to authority book, course, etc.).

___ Other.

6a. Do you think the student's fluency was low?

Yes No

b. How do you know?

___ Reasoned in terms of relationships (cause/effect--pupils who have this problem do this [behavior]).

___ Reasoned in terms of experience with other students.

___ Reasoned in terms of authority (general reference to authority book, course, etc.).

___ Other.

APPENDIX D

CLINICIAN RESPONSES ON FORMAL AND INFORMAL MEASURES

DATA REQUESTED AND NOT AVAILABLE FOR FOUR CASES

Subject Number	Case Number	Run Number	Data Requested, Not Available
101	3	1	None
101	4	2	None
101	3	3	None
102	2	1	Ekwall Phonics Inventory or Woodcock or Gates-McKillup
102	1	2	None
102	2	3	Gates-McKillup Nonsense Word List Bader's Phonics Test
103	1	1	None
103	3	2	Psychological report ITPA
103	1	3	Iowa or Stanford Phonetic analysis for Durrell Gates-McKillup, sounds and their relationships
104	4	1	None
104	2	2	None
104	4	3	None
105	3	1	None
105	4	2	None
105	3	3	Botel Dolch Word List Memory Battery of Woodcock Botel (2nd request) Creative Writing ITPA Speech report Counselor report Kottmeyer Miscue Inventory

Subject Number	Case Number	Run Number	Data Requested, Not Available
106	4	1	Detroit or visual perception
106	2	2	Auditory discrimination Health record Gates-McKillup oral reading and examiner's comments Cover sheet of Durrell or summary profile Durrell Spelling or Gates-McKillup Weschler full-page profile
106	4	3	Gates-McKillup
107	1	1	Stanford Achievement Durrell Spelling Test Handwriting sample
107	3	2	Peabody vocabulary and comprehension Motor ability test Informal Reading Inventory
107	1	3	Peabody score interpretation Wepman auditory blending
108	2	1	Psychological tests Informal reading inventory Kottmeyer ITPA or auditory memory Expressive language indication
108	1	2	Visual acuity Visual discrimination Left-right dominance or laterality Writing sample or writing sample on spelling test of Durrell Drawing test or Bender-Gestalt Dictated sentences Visual/motor coordination skills Wepman Reversals test
108	2	3	Examiner's comments for Individual Reading Analysis Durrell Visual Memory, intermediate level

WRITE-IN DIAGNOSTIC STATEMENTS FOR "OTHERS" CATEGORIES OF READING DIAGNOSTIC CHECK LIST

Subject Number	Case Number	Run Number	Category Number	Value (S,W,O)	Verbatim Comment for "Others" Diagnostic Statements
101	3	1	IV	W	Big words, little words, and little print gave him trouble.
				W	Omits sounds.
				W	Adds sounds.
101	4	2	V	W	Did not attempt to decode words phonetically.
				S	Reliance on context helps comprehension.
			VI	W	Examiner had to pronounce unknown words.
				W	Had only one correct word on the 3rd grade list.
			X	W	Lack of basic reading skills.
				W	Posture was poor as he kept his eyes 5 inches from the paper.
101	3	3	III	W	Teacher had misconceptions about the relationships between interests and actual school achievements.
				W	Gates-McGinitie, he scored 6.5 grade equivalent in vocabulary and 5.3 grade equivalent in comprehension.
			VI	W	Auditory weaknesses were reflected in the way he recognized words in isolation.
			VII	W	Omitted, added, and wrote incorrect sounds.
				W	Errors generally sensible
				W	It should be noted that he tried all lists.

Subject Number	Case Number	Run Number	Category Number	Value (S,W,O)	Verbatim Comment for "Others" Diagnostic Statements
101 (cont'd)	3	3	VIII	W	Was probably working on word recognition.
			IX	W	Listening comprehension.
102	2	1	IX	S	Listening comprehension.
102	1	2	VII	W	"le" ending.
102	2	3	VII	W	Word endings.
			IX	O	Listening comprehension.
103	1	1	IX	S	Semantics.
				S	Syntax.
103	3	2			None.
103	1	3			None.
104	4	1			None.
104	2	2			None.
104	4	3			None.
105	3	1	X	W	Spelling.
105	4	2			None.

Subject Number	Case Number	Run Number	Category Number	Value (S,W,O)	Verbatim Comment for "Others" Diagnostic Statements
105	3	3	VI	W	Words in print, not oral.
			IX	W	Influence of word recognition on comprehension.
				W	Influence of word miscue on comprehension.
			X	O	Spelling achievement.
				O	Handwriting achievement.
106	4	1	IX	S	Listening comprehension.
106	2	2	I	O	Frequent absences due to health.
			II	W	Application of concepts (ability in).
			IV	W	Auditory figure ground related to performance in academics.
				S	Spatial relationships general.
106	4	3	IV		No comments.
107	1	1	V	W	Miscue unacceptable.
107	3	2	I	W	Vision-muscle imbalance.
			V	W	Oral reading--poor phrasing.
				W	Oral reading--repetitions.
			X	W	Spelling--omits sounds, adds sounds, incorrect sounds, not grade level

Subject Number	Case Number	Run Number	Category Number	Value (S,W,O)	Verbatim Comment for "Others" Diagnostic Statements
107	1	3	V	W	Miscue contextually unacceptable.
			IX	W	Comprehension--literal.
108	2	1	II	W	Relation of listening, achievement to mental capacity.
108	1	1			No comment.
108	2	3	I	O	Respiratory complication.
			V	W	Oral reading repetition.
			X	O	Learning disabled profile.

UNEDITED DIALOGUE OF INFORMAL DEBRIEFING FOR SUBJECT 104, RUN 3, CASE 4

E: Before we end this session I have some questions I'd like to ask you regarding your opinions about some aspects of reading diagnosis.

1. In your opinion, what kinds of information should be included in a "good" diagnosis?

C: Well, the first thing that I'm concerned about is to make sure that there are no unusual factors such as hearing deficit or hard of hearing or partially sighted or some emotional factors that I'm not aware of such as child abuse or some other factor which is affecting the child either physically or emotionally. But most of the kids that I see that are in the regular school rooms, the so-called typical kid, most of them come from so-called normal environments and they're not wearing glasses and they don't appear to be hard of hearing. Once that's out of the way, I kind of discount any emotional effect on the child or any abnormal physical effects on the child. The first thing I want to know about the kid is what does he know about words; how many words does he know; and I usually check this out by--I usually have kind of an informal--

I pick up something quite easy for them to read, have something they want to read to me and the kid normally picks up something that he can read. Occasionally he'll pick something quite difficult, then I have to kind of gear them toward something else. After having them read to me orally, I get kind of a hunch where they are already but then I give them the Slosson test, the Slosson Oral Reading Test. You get some kind of an approximate grade score or an idea of how many words they know by sight. Then I usually follow that up with some type of inventory. I like to use the Ekwall because it has some nouns in it, other than just the verbs or other words that Dolch has, to see which words they miscue on when they read words in isolation so I can see if there are some consistencies in their miscues. I also find out which words they know. That's going to be helpful to me later on when I start working with them in remediation. Taking words that they know for example that the regular words that they follow some type of word family. I might be able to show them some phonic skills in the area of word families or whatever I feel they're weak in--blends, digraphs, or word endings. I take the words they already know and try to work out some word patterns or some systematic approach. So the kid says, "Yeah, there is some regularity to our language." After getting an idea of what they know about words in isolation and how they attack words in isolation that they don't know, that they don't know instantly, I get some kind of a clue into their understanding of the phonic system, and I usually try to administer either an informal or a phonic or a more formal phonics test to see if they know initial and final consonants, blends, diagraphs, multisyllabic words, vowel rules. I usually don't get too carried away with it. The kids I work with are usually in about 4-5th grade, and they're down reading about 2nd, 3rd grade level. About the only thing I'm really concerned about in the area of vowels is do they understand the word families and short vowels, and do they understand the vowel rules that have high frequency of application of which I think there are very, very few. Then I'm kind of concerned about what do they do with the knowledge of the words that they have, instant sight word recognition and their knowledge of phonics. How do they apply this? Are they still word-by-word plodders? I've run into some kids that know a lot of words or they'll know 200 Dolch words or 210. They'll have a pretty good understanding of the phonics rules, they can sound out most any word that they see in isolation and yet when they--when you ask them to read it's almost word-by-word plodding. So, I go to the oral reading to see how they apply this if they do read in phrases. Do they read, "under - the - " or "under the." It's got to be "under the something." And some of the kids just won't do that. Or it'll be "under the bridge of the cow," and they don't apply our punctuation system. So kids that know a lot of basic sight words and kids that know the phonics system, but don't apply it are really kind of confusing because they give you the appearance of being able to read fluently. If you don't listen to them read orally you don't really have an understanding that they are reading word-by-word. Particularly in the upper grades some of the kids can survive with this word-by-word reading. And they can still get satisfactory scores so the teacher doesn't get too upset. And there's not a lot of oral reading going on in the upper grades. But I think that the oral reading is the method that we should use for diagnosis to find out, "Hey, what is the kid doing right and what's he doing wrong?" Where I find a lot of breakdown is in the application of systems. Either they don't know the systems or even if they know, if they know the systems, they don't apply them in their reading. I don't get too carried away with comprehension. Most kids that I work with that are deficient in phonics and structural analysis skills and basic sight word knowledge, they still can understand anything written at their grade level. The stories aren't that complicated and if they can read it they can usually give you back the factual, literal recall. I still think we're really shaky on what comprehension is other than the literal recall. I think we're shaky in the cognitive understanding of what happens in the reading act. There are some basic comprehension skills that I assess other than literal recall, but most of the time I don't get too carried away with it because I find the kids will understand what they're reading if they only know the system and apply the systems. Because most of the kids understand if you read to them. That's why I like to give this listening comprehension. For example, the kid we worked on today, oral and silent reading--they mirrored one another. We couldn't hear his silent reading but I suspect that he read silently just like he read orally, word-by-word, made the same miscues on those high frequency words "there" for "three," "then" for "when" and "when" for "then" on different occasions. But yet, when he was read to, when you asked him for literal recall, he understands. Kids understand. In fact, most of them are quite sophisticated in their understanding of the spoken words. They understand what they hear on television. They understand what their friends say to them. But if what was said on television was reduced to a script, such as it was in this Walton's Pearl Harbor thing, and there was one other one that they did, once you put it down visually, plop it in front of them, their comprehension goes all to hell; they don't know the systems; they don't know the words; they don't know the phonics, or they just don't put it together. Visually they just get nothing. Yet when they see it on T.V. and you ask them the questions about it, they've got it. So, I don't get too carried away with comprehension. I think the kids in the elementary grades, the normal kid, doesn't have any trouble comprehending the material that he's reading.

E: I think you've kind of answered the next question, which is:

2. How do you usually go about a diagnosis?

Did you finish that? Let's see--we got to the informal.

C: Yeah, the informal. I check sight words. I check phonics. I check structural analysis, and then I check application of the systems and then I do some limited assessment in comprehension.

E: And you feel that that's pretty much in answer to the first question too, like what do you think a good diagnosis should include?

C: That's what I think it should include.

E:

3a. Thinking in terms of the case with which you just worked, do you think that _____'s instant word recognition skills were low?

C: Yes, definitely.

E: And how did you know that?

C: By his scores on the Slosson. I think it was 42, which indicated about the beginning 2nd grade.

E:

3b. What about _____'s word analysis skills?

C: Very poor because he has just a confused, hodge-podge understanding of phonics. He seems to know initial consonants, but he doesn't look at word endings. His miscues are--well, you could find probably half a dozen patterns of miscues. He just doesn't seem to do anything right other than maybe clue in on that initial consonant. Sometimes he doesn't even do that when he says "saw" for "was" or "yes" for "say."

3c. Do you think that _____'s comprehension skills were low?

C: Oral and silent reading. And the reason they were low in oral and silent is because the kid can read, can't apply the systems, doesn't know the words, and doesn't know the phonics. But when he was read to with the Durrell, he scored very well up to 5th grade, answered 7 of 9 on the 5th grade one.

E: So you're saying his listening comprehension is good?

C: I could read to him 5th grade material. He understands it.

E: So, you've determined that the oral and silent were low and that the listening was ok. How did you know this?

C: By the scores and the reason I like to get these scores. I think that difference between listening comprehension and silent--oral in his case which were 2 years apart, this gives you an idea of where you can bring that kid; his potential. Let's say he got 9 out of 9 at the 7th grade, then that indicates that the kid's really got a lot on the ball; something's screwed up. The kid understands up to 7th grade, who's only reading on 2nd grade level. If we can only straighten out where his problem is we should be able to bring him up to 7th grade level. On one of the other children that we had when we did the other ones. Sometimes we get a little confused when we say well how far can we bring this kid along. We find that their oral reading comprehension, their silent reading comprehension, their listening comprehension are about all the same. Then I start saying, "Hey, maybe the kid's working up to grade level even though he's in the 4th grade reading at 2nd grade level. Maybe we shouldn't be pushing this kid any harder."

E: That's--but when you see the big discrepancies. . . .

C: You see the big <u>discrepancy</u>. This gives you an idea of where you can bring the kid to his potential.

E:

3d. What about Brian's reading fluency? Do you think that was low?

C: Yes, very definitely and it goes back to the application of the systems. He doesn't understand the systems, so we can't expect him to apply the systems. He's just a word-by-word plodder. If he miscues the words in isolation, he's still going to miscue them in his oral reading. Although he made attempts, he realized that "Hey, this just does'nt make sense if I say 'then' instead of 'when'" and twice in his oral reading, after he completed a whole sentence, he went back knowing that "Hey, it can't be. It didn't make sense." So he went back and plugged in the right word. Oh, he said "next to" for "near." Well, "next to" and "near" are pretty close in meaning. So, he knows that reading has to make sense, and he's bothered when it doesn't make sense. He just doesn't continue right on. There was one time, he kind of went on because it didn't make any sense. But you can almost sense in his reading when he miscued and there was a loss, a real loss in comprehension, he went back and reread the whole sentence.

E: Well, do you have any other thoughts concerning reading diagnosis and remediation that you'd like to share with me? Just kind of off the cuff, any concerns you have?

C: I am a little. The thing like this kid we did today, you know, strong auditory learner in his spelling of nonsense words--"car-plite." You could see him putting in the "c" and the "r" and the "p" and the "i"--maybe mixing them up because he doesn't apply it. But that kid heard the words and could find them on the hearing sounds in words. He got 28 out of 29. When you show him the word visually, he doesn't know how to look at that word. He just--I think he scored 1.5 on that. So, when he sees something, he has all kinds of trouble. But with the combination of hearing and seeing I think that kid can be remediated. He can be helped a lot.

E: And along that line, when you say you think he can be remediated, and thinking maybe more in terms of college students which you might be teaching, do you have any suggestions for improvement of instructing reading teachers, maybe thinking in terms of diagnosis and remediation?

C: Yeah, one of the things that I think that, first of all, some of the courses that I've taught other than 830C, the beginning course. They don't seem to know the scope and the sequence of phonics skills.

E: The college students?

C: The college students taking the graduate level reading courses, they don't seem to know the scope and sequencing of phonics skills and the other thing that they seem to be weak in is when they look at the words in isolation and they're looking for error patterns they seem to have trouble seeing that "Hey, the kid's got a vowel problem" or that the kid is miscuing on words that are highly similar in appearance. They'll look at the words "there" and "three" and say that the kid doesn't know his vowels. Well, when I look at "there" and "three" t-h-e-r-e and t-h-r-e-e, I call those highly similar appearing words and that's not a vowel problem. "Want" and "went" they'll say is a vowel problem. Those are both, well "want" in particular being an irregular word, it isn't a vowel problem if a kid doesn't know "want" when he sees it he says "went."

E: How might you suggest that we improve this in teaching?

C: Oh, I could talk for a long time. There are the things I see that teachers do.

E: Maybe you should light on one that bothers you the most.

C: Well, another thing that does bother me is we spend so much time on words in isolation and working on helping kids read words in isolation. They so infrequently read words in isolation. I think we have to tie our instruction in phonic skills, in structural analysis skills and teaching of basic sight words, we could be tieing this back into the context. In the beginning, context with noun phrases, verb phrases, prepositional phrases, and as soon as possible sticking it into whole sentence context. I see too much instruction on an isolated word basis and like light I think for example, the word light, he gives you light, doesn't give you a hell of a lot out of isolation or out of context. It can mean so many different things. Or words like "want," "need," or whatever. Pick any of those isolated sight words. They don't mean much out of isolation. So I think we should be making an attempt to teach the phonics skills, the structural analysis skills, and basic sight words in context as soon as possible because then this is going to improve the fluency and is going to get the kid to applying the system, the very systems that belong to him.

E: How do you see diagnosis and remediation related or interrelated?

C: It should be a constant, ongoing process. How you make your initial diagnosis, and you start working on it, you're going to find out other things that you missed in your initial diagnosis and then as you're working with the kids and you see progress in various areas, you kind of start working on the strengths and concentrating on the deficits where the kids really have problems, but the diagnosis and remediation have to be continual then.

E: Do you think the diagnosis is important as such? You know, well, let's put it this way. Do you see yourself being able to successfully remediate a student without going to a formal kind of diagnosis, maybe the one that you mentioned as being a good diagnosis?

C: Well, the one I talked about I think is a simple, well maybe a little bit too simplistic or too simplified, but just because of the problem of time and number of students I have to get realistic and maybe a little bit pragmatic about how I diagnose the kids. But I think you have, you know, you have to have an initial diagnosis. If you don't know where the kids are, you don't know where the hell you're going. Poor plan's better than no plan at all. So, as soon as possible in the first grade like at the grade level I teach, as soon as the kids get in there, I find out which letters they know, which letters they confuse, which sounds they know, do they know initial consonants, are these kids already reading, what words do they know, what is their knowledge of basic sight words, and how much do they know about phonics, and take off from there. But, if I just walk in and you know and assume that the kids know such and such and started teaching in different groups or something there would be just a hodge-podge.

E: Do you think that the graduate students that you teach have an idea of how to go about a good diagnosis and remediation? Do you think they have a model?

C: I think they will after the seminar.

E: So you do work toward helping them develop a pattern of a good diagnosis and what's included in it--the diagnosis?

C: Yeah, I try to do it again in a practical way. They don't have time to sit down and write a 3-hour diagnosis. In one diagnosis class that I worked with, they did one full-blown 10-12-14-16 page diagnosis on their most difficult child and then in the last 2 days we sat down and did it in 15 minutes. I'd say, "Now in 15 minutes tell me what the kid knows about the sight words. Tell me about his flash recognition and his analysis of basic sight words, look at your Durrell, look at your SORT, look at your Ekwall list and your Dolch list and for 15 minutes tell me what the kid does right, tell me what he does wrong." And then we'd go into the phonics. I'd say, "All right, hey for the next 20 minutes you're going to say, 'Hey, what does the kid know about phonics?' Tell me about initial consonants, final consonants, blends, diagraphs, vowels, suffixes, prefixes. Put it down. You've got 20 minutes, do it. Then tell the teacher next door who's got the other 2nd grade classroom about it." Then in another 15 minutes I'd say, "What does the kid know about word patterns? What does he know about punctuation? What does he do with noun phrases, verb phrases, prepositional phrases? How does he read when he reads orally? Write it in 15 minutes." And then, finally, I say, "OK, now comes the toughy. Tell me about comprehension. What does the kid do with comprehension? How is comprehension affected by his knowledge of phonics, sight words, his application system?

What is the student's comprehension? Write for 15 minutes." So, in four 15-minute periods they've pretty well covered some type of a--they draw some theoretical ground issues in those four ideas. This is what I hope they do when they get back to the classroom.

E: Well, this leads me to my final question, which is how did you feel about the task that you had to do for the three sessions and what about the time pressure that you were under? Did you feel it was realistic or difficult?

C: What did we do? We had how many minutes to diagnose?

E: Forty-five minutes to request information on the case and then 25 minutes to write the diagnosis and 25 minutes for remediation.

C: Yeah. That's realistic. It may be longer than the people out in the field would get in a class of 40 or maybe a class of 50. You've got 50 kids to work with in your caseload, you may not have that long.

E: Now we were, I guess, most concerned about the time of the actual writing, but since it was a time-press situation, we wanted to see what people could do with limited time. Did you feel that you improved any in your writing between the first and last time?

C: Oh, just for example I think the others I had 13-14 diagnostic comments that I had to identify and record in the book. Today I think I had 21. I think I performed better today than I did the other two times and that could be familiarity with what was expected of me--the process that I was going to be going through and less anxiety.

E: Well, I certainly appreciate your time. This concludes our three clinic sessions.

Table D1.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 101, Case 3.

What information should be included in a "good" diagnosis? (1)	How do *you* go about a diagnosis? (2)
1. A look at behavior.	1. Get initial information about child from his teacher and his parents; also speech teacher or social worker if that applies.
2. Recognition of words in isolation. Pre-reading skills strengths and weaknesses.	2. I might also observe child in classroom.
3. Ability to figure out words in reading situation--oral, silent and listening.	3. Then I interview the child, get acquainted and get self-report information.
4. Oral reading for clue to comprehension by way he figures out words.	4. I use a word list or quick measure of reading behavior and grade ability; sometimes I use child's own book.

Table D2.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 102, Case 2.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. Instant word recognition.	1. Home background through telephone interview with mother.
2. Word-attack strategies.	2. Slosson Oral Reading Test.
3. Fluency.	3. Parts of the Durrell
4. Rate.	4. Informal reading inventory for comprehension.
5. Literal comprehension.	5. I.Q. test only if appears to be a problem.

Table D3.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 103, Case 1.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. It is based on the individual child. But I guess I ask for information on the environment and maybe I.Q.	I look at:
2. Information most helpful is looking at the child's oral reading performance. Oral reading gives hunches related to ability to attack words--attempt to decode words that are not known. If he performs well there I wouldn't need to go to Dolch Word List or Slosson.	1. Word recognition
3. Should look at ability to use sound-symbol relationships.	2. Word analysis
4. Comprehension, including use of context clues.	3. Comprehension, and
5. Standardized test scores for grade placement.	4. Attitude.
6. Background information.	
7. Why child referred.	

Table D4.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 105, Case 3.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. Focus on strengths of child.	1. Intensive interest inventory to get acquainted.
2. Relate field of reading to child's other language arts skills.	2. Screening for quick evidence of grade level and for basic skills--SORT, Slosson, I.Q., Botel word opposites.
3. Assurance on elementary skills, especially older children, such as basic Dolch sight words, sounds of letters or blends, word analysis.	3. Gates-MacGinitie to confirm reading level.
4. What motivates a child--the affective process.	4. Look for divergence that would respond to whatever I begin to find out about child.
5. Self-diagnosis by the child.	

Table D5.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 107, Case 1.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. You need silent and oral reading and compare the two.	The way I just explained it.
2. You should use listening comprehension or if there's a WISC available that would do for potential.	
3. I use the Slosson with older kids or Dolch with younger ones for sight word recognition and grade level score.	
4. Most important thing is listening to reading.	
5. An informal inventory to support a standardized test or for a quick check a group silent test.	
6. An interview with student.	
7. Background information from records, talk with teacher.	

Table D6.--Subject responses to informal questions regarding memory (1) and strategy (2)--Subject 108, Case 2.

What information should be included in a "good" diagnosis? (1)	How do you go about a diagnosis? (2)
1. Listening comprehension test if not an I.Q. measure. 2. Extensive informal reading inventory--oral and silent paragraphs with comprehension checks and word recognition skills. This yields information as to child's performance on school-type tasks, tests, etc. 3. Depending on case, you would go into individual analysis of either specific reading skills or perhaps auditory areas or visual areas. Efficient diagnosis should eliminate testing in areas such as visual unless indication of a problem.	It depends on the situation. If it's to confirm a teacher's finding I would use just the reading portion of a diagnosis. If I have the time I prefer a more thorough diagnosis that gives the advantage of checking visual and/or auditory difficulties and some measure of capacity and strength.

Table D7.--Subject responses to informal questions regarding specific aspects of reading--Subject 103, Case 1.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other Response: Looking at Gates-MacGinitie, his performance was about a year or more below his grade placement. Another clue was Dolch Word performance; he could only identify 61% of 1st grade words. Third, he read with a number of hesitations.	__ Yes ✓ No	__ Model __ Experience ✓ Authority __ Other Response: On word list he did know his initial consonant sounds, his short vowels and short vowel phonograms. This indicated to me he was doing pretty well for a third grader.

3a. Was reading comprehension low?	3b. How do you know?	4a. Was reading fluency low?	4b. How do you know?
__ Yes ✓ No	__ Model __ Experience ✓ Authority __ Other Response: Because of poor word recognition skills.	✓ Yes __ No Response: Poor.	✓ Model __ Experience __ Authority __ Other Response: There were many hesitations when he was reading. His sight word vocabulary is inadequate so you know right away there's a fluency problem in reading.

Table D8.--Subject responses to informal questions regarding specific aspects of reading--Subject 107, Case 1.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other
	<u>Response</u>: When he was reading he called "kitten" "chicken." He did not stop when it didn't make sense. A chicken doesn't drink milk. There were other errors that I picked up again on the Dolch.		<u>Response</u>: He doesn't take a word apart; he uses general word parts. He also had some reversals and ignored medial sounds.

Table D8.--Continued.

3a. Was reading comprehension low?	3b. How do you know?	4a. Was reading fluency low?	4b. How do you know?
__ Yes ✓ No	✓ Model __ Experience __ Authority __ Other	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other
Response: Surprisingly that was really very good.	Response: All the way through the oral reading it was exceptionally good even though he didn't always make the best sense--he would work around it. Some of his miscues were way off, but they made sense and were acceptable. His silent reading comprehension wasn't as good as the oral, so that would tell me that he needs reinforcement of hearing what he's reading.	Response: It was bad.	Response: It's almost word-by-word. He wasn't always searching for a word; he just hasn't developed fluency. He also ignored punctuation and some of his phrasing was poor.

Table D9.--Subject responses to informal questions regarding specific aspects of reading--Subject 102, Case 2.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
__ Yes ✓ No	✓ Model __ Experience __ Authority __ Other	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other
Response: His was fair.	Response: From Slosson Oral Reading Test; he gets through the first few lists without making very many mistakes. Fact he didn't go much farther was because he didn't know phonics skills.		Response: He doesn't attempt to analyze words; would say he didn't know or would guess, sometimes couldn't get first letter.
3a. Was reading comprehension low?	**3b. How do you know?**	**4a. Was reading fluency low?**	**4b. How do you know?**
✓ Yes ✓ No	__ Model __ Experience ✓ Authority __ Other	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other
Response: Silent low; oral and listening not as low.	Response: Oral and silent reading paragraphs.		Response: Pauses between words, does not observe punctuation; phrasing better than expected.

Table D10.--Subject responses to informal questions regarding specific aspects of reading--Subject 108, Case 2.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other	✓ Yes ✓ No	__ Model __ Experience __ Authority ✓ Other
	Response: He had miscues in both the word recognition test and in the oral reading.	Response: They were low for his grade level but not low for his overall capacity for reading.	Response: He showed a tendency to be able to analyze words with some encouragement and assistance.

Table D10.--Continued.

3a. Was reading comprehension low?	3b. How do you know?	4a. Was reading fluency low?	4b. How do you know?
✓ Yes ✓ No	✓ Model __ Experience __ Authority __ Other	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other
<u>Response</u>: Reading comprehension was above his reading efficiency skills but low for his age level. I suspect it was consistent with his mental capacity or whatever factors might have influenced that.	<u>Response</u>: His verbal performance was depressed in general on capacity tests, and I suspect that's related to some auditory deficiencies. That's a complicated factor and could be interpreted as possibly affecting the lower intelligence score, lower language skills, which in turn would indicate that you would expect lower comprehension.		<u>Response</u>: By listening to him read, it appears he doesn't read very often. As I tell my kids, "he sounds like a rusty, old tractor that's been out in the snow all winter."

Table D11.--Subject responses to informal questions regarding specific aspects of reading--Subject 101, Case 3.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
__ Yes ✓ No	__ Model __ Experience __ Authority ✓ Other	✓ Yes __ No	✓ Model __ Experience __ Authority __ Other
	Response: Not that low--in 7th & scores 6.8 on instant word recognition. But school district has reputation for having children doing well and so many of his peers are probably above him. He's conceiving himself as being poor, so in that case he's probably low.		Response: I'm almost positive that he doesn't have word recognition skills because of a hearing problem. He can't divide words into syllables. Auditory would be helpful in word analysis and he couldn't do the auditory things.

Table D11.--Continued.

3a. Was reading comprehension low?	3b. How do you know?	4a. Was reading fluency low?	4b. How do you know?
__ Yes ✓ No (Qualified)	✓ Model __ Experience __ Authority __ Other Response: It was low according to the grade level achievement scores that we have but for him and the way he was recognizing words, he did okay.	✓ Yes __ No	✓ Model __ Experience __ Authority __ Other Response: In oral reading it wasn't very good; a lot of hesitations, and repetitions. He had to stop and correct himself and I'm wondering if that wasn't insecurity. He couldn't analyze all those words--had to use context a lot.

Table D12.--Subject responses to informal questions regarding specific aspects of reading--Subject 105, Case 3.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other	✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other
Response: Probably.	Response: He was inconsistent--word "gradual"; he got it correct twice and missed it once.	Response: Probably.	Response: Material from Gates-McKillup, Durrell & Ekwall.
3a. Was reading comprehension low?	**3b. How do you know?**	**4a. Was reading fluency low?**	**4b. How do you know?**
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other	✓ Yes __ No	✓ Model __ Experience __ Authority __ Other
	Response: Durrell Oral Reading low in comprehension and was confirmed particularly through Gates MacGinitie comprehension.		Response: Phrasing so terrible it would destroy fluency and comprehension (tape of oral reading).

Table D13.--Subject responses to informal questions regarding specific aspects of reading--Subject 106, Case 4.

1a. Was instant word recognition low?	1b. How do you know?	2a. Were word-analysis skills low?	2b. How do you know?
✓ Yes __ No	__ Model __ Experience ✓ Authority __ Other Response: He's 4th grade and Dolch and Durrell came out at high 2nd grade, which is low for 4th grader.	✓ Yes __ No	__ Model __ Experience __ Authority ✓ Other Response: He used initial consonant blends pretty consistently but other than that I didn't see anything he used consistently. Once in awhile it looked like he used a final consonant clue; a blend here and there, but really extremely limited knowledge of vowels or variance of word parts or structural analysis.

Table D13.--Continued.

3a. Was reading comprehension low?	3b. How do you know?	4a. Was reading fluency low?	4b. How do you know?
✓ Yes __ No	✓ Model __ Experience __ Authority __ Other Response: Reading comprehension better than reading skills. He was able to get enough information from the key words and phrases and so forth that he was able to read that he could comprehend pretty well on that; this based on Durrell reading and listening.	✓ Yes __ No	✓ Model __ Experience __ Authority __ Other Response: From teacher's and examiner's comments that he was a slow reader, but I would suspect that because of his low decoding skills.

APPENDIX E

EXAMINER'S NOTES FORM AND LETTER TO SUBJECTS

EXAMINER'S NOTES

Clinician's Name: ______________________ File No. __________

Case Name: _______________

Date: ___________________

Time Session Began: ______________

Time Session Ended: ________________

Total Time of Session: ______________

Time Initial Contact Began: __________

Time Initial Contact Ended: __________

Total Time Initial Contact: __________

Time of First Cue Request: ___________

Time Cue Interaction Ended: __________

Total Time on Cues: ________________

Time Written Dx Began: ______________

Time Written Dx Finished: ___________

Total Time for Written Dx: ___________

Time Rx Began: ____________________

Time Rx Finished: _________________

Total Time Rx: ____________________

Time Dx Transfer Began: _____________

Time Dx Transfer Ended: _____________

Total Time for Dx Transfer: __________

Time Hx/Ob Check List Began: _________

Time Hx/Ob Check List Completed: ______

Total Time on Hx/Ob Check List: _______

Allowed Cue Time
Began:
+ 25
Remind:
+ 20
End:

Allowed Dx Time
Began:
+ 25
End:

Allowed Rx Time
Began:
+ 25
End:

Clinician: ________________

Case Name: ________________

Date: ____________________

Problems with session:

Clinician's reactions:

Other:

MICHIGAN STATE UNIVERSITY

INSTITUTE FOR RESEARCH ON TEACHING
COLLEGE OF EDUCATION · ERICKSON HALL

EAST LANSING · MICHIGAN · 48824

November 29, 1978

Dear

Thank you again for agreeing to be a clinician in the reading research project for the Institute for Research on Teaching (IRT). As I discussed with you at our final session, I need to have you transfer each of your written diagnoses to the Diagnostic Check List for a second time. The purpose of having you repeat this task is to check the reliability of the Diagnostic Check List.

Enclosed are copies of your three written diagnoses along with three Diagnostic Check Lists and a stimulated recall sheet. When you have completed all three transfers of your diagnoses to the check lists, please place all of the materials you were sent in the enclosed envelope and mail them promptly.

Please follow these instructions:

1. Please do not rush with this task. You may want to work on each diagnosis on a different day.

2. Begin with your written diagnosis for case number 1, __________. Do not look at the other diagnoses.

3. Locate and circle all of the diagnostic or reading factor statements in your written diagnosis.
(Note: A diagnostic statement is the identification of a factor or variable which helps determine the state or condition of a student's reading performance. It may be one sentence or several and it may also just be part of a sentence. In addition, the diagnostic statement may be a strength, a weakness or just an observation.)

4. Circle and number the diagnostic statements or factors.

5. Locate on the cover sheet of the Diagnostic Check List the main category into which your diagnostic statement would most likely fit.

Page 2
November 29, 1978

(Note: Space is provided within each main category for "Other Statements". You may copy any statements from the written diagnosis under "Other Statements" if you are able to determine the main category but are unable to find a matching statement on the check list in that same category. However, statements in the written diagnosis for which you can not determine a category should be copied under the main category "Other" in the check list. Please make as limited use as possible of the statements and the category called "Other".)

6. Find that main category in the check list and locate a statement within it which best matches your circled statement.

7. Decide if your statement indicates a strength, a weakness or is just an observation which you made; then place the number of your circled statement under one of the column headings.

8. If a stimulated recall sheet is included with your materials, circle the diagnostic statements in the stimulated recall as you did in the written diagnosis but use letters A, B, C, etc. rather than numbers to designate each statement. Next transfer the letters to the Diagnostic Check List as you did for the statements in the written diagnosis.

9. Check to make sure that you did not omit any statements from the check list.

10. Repeat procedures 2-8 for case number 2,_____.

11. Repeat procedures 2-8 for case number 3,_____.

12. Date each check list for the date you complete it.

13. Please be prompt in completing the check lists and in returning them.

In the interest of protecting the work of the IRT, please do not discuss the nature of the tasks you have been doing with others in the field of reading.

You should be receiving a check very soon for your work on this project. If you do not, please call me collect at (517) 332-2351.

On behalf of the Institute for Research on Teaching, thank you for participating in this research project.

Yours very truly,

Lynne Hoffmeyer

Lynne Hoffmeyer
College of Education
Michigan State University

Enclosures:
3 Diagnostic Check Lists
3 Diagnoses
3 Stimulated Recalls (if you used them)
1 Envelope for return mailing

APPENDIX F

CUE DOMAIN SAMPLE FOR A CASE

CUE DOMAIN SAMPLE FOR A CASE

1	BKG 2	Background Information--Biographical Data--EC
2	BKG 8	Background Information--Physical/Health--EC
3	BKG 14	Background Information-- /Family--EC
4	BKG 20	Background Information--Classroom Information--EC
5	DOL 1	Dolch Basic Sight Vocabulary--TS
6	DOL 2	Dolch Basic Sight Vocabulary--EC
7	DOL 3	Dolch Basic Sight Vocabulary--TB
8	DOL 4	Dolch Basic Sight Vocabulary--AR
9	DOL 5	Dolch Basic Sight Vocabulary--TD
10	DUR 1	Durrell Oral Reading--TS
11	DUR 2	Durrell Oral Reading--EC
12	DUR 3	Durrell Oral Reading--TB
13	DUR 4	Durrell Oral Reading--AR
14	DUR 5	Durrell Oral Reading--TD
15	DUR 7	Durrell Silent Reading--TS
16	DUR 8	Durrell Silent Reading--EC
17	DUR 9	Durrell Silent Reading--TB
18	DUR 11	Durrell Silent Reading--TD
19	DUR 13	Durrell Listening--TS
20	DUR 14	Durrell Listening--EC
21	DUR 15	Durrell Listening--TB
22	DUR 17	Durrell Listening--TD
23	DUR 19	Durrell Word Recognition and Word Analysis--TS
24	DUR 20	Durrell Word Recognition and Word Analysis--EC
25	DUR 21	Durrell Word Recognition and Word Analysis--TB
26	DUR 22	Durrell Word Recognition and Word Analysis--AR
27	DUR 23	Durrell Word Recognition and Word Analysis--TD
28	DUR 25	Durrell Visual Memory of Words-Primary--TS
29	DUR 26	Durrell Visual Memory of Words-Primary--EC
30	DUR 27	Durrell Visual Memory of Words-Primary--TB
31	DUR 29	Durrell Visual Memory of Words-Primary--AR
32	DUR 31	Durrell Hearing Sounds in Words--TD
33	DUR 32	Durrell Hearing Sounds in Words--EC
34	DUR 33	Durrell Hearing Sounds in Words--TB
35	DUR 35	Durrell Hearing Sounds in Words--TD
36	DUR 37	Durrell Sounds of Letters--TS
37	DUR 38	Durrell Sounds of Letters--EC
38	DUR 39	Durrell Sounds of Letters--TB
39	EKW 1	Ekwall Phonics Survey--TS
40	EKW 2	Ekwall Phonics Survey--EC
41	EKW 3	Ekwall Phonics Survey--TB
42	EKW 5	Ekwall Phonics Survey--TD
43	GMG 1	Gates-MacGinitie-Vocabulary--TS
44	GMG 2	Gates-MacGinitie-Vocabulary--EC
45	GMG 3	Gates-MacGinitie-Vocabulary--TB
46	GMG 7	Gates-MacGinitie-Comprehension--TS
47	GMG 8	Gates-MacGinitie-Comprehension--EC
48	GMG 9	Gates-MacGinitie-Comprehension--TB

49	GMG 13	Gates-MacGinitie-Speec & Accuracy--TS
50	GMG 15	Gates-MacGinitie-Speed & Accuracy--TB
51	GMK 1	Gates-McKillup Recognizing & Blending Common Word Parts--TS
52	GMK 2	Gates-McKillup Recognizing & Blending Common Word Parts--EC
53	GMK 3	Gates-McKillup Recognizing & Blending Common Word Parts--TB
54	GMK 4	Gates-McKillup Recognizing & Blending Common Word Parts--AR
55	GMK 5	Gates-McKillup Recognizing & Blending Common Word Parts--TD
56	INF 2	Informal Oral Reading--EC
57	INF 3	Informal Oral Reading--TB
58	INF 4	Informal Oral Reading--AR
59	INF 5	Informal Oral Reading--TD
60	IRA 3	Individual Reading Analysis--TB
61	IRA 5	Individual Reading Analysis--TD
62	PEA 1	Peabody-Reading Recognition--TS
63	PEA 2	Peabody-Reading Recognition--EC
64	PEA 3	Peabody-Reading Recognition--TB
65	PEA 5	Peabody-Reading Recognition--TD
66	PEA 7	Peabody-Reading Comprehension--TS
67	PEA 8	Peabody-Reading Comprehension--EC
68	PEA 9	Peabody-Reading Comprehension--TB
69	PEA 13	Peabody-Spelling--TS
70	PEA 14	Peabody-Spelling--EC
71	PEA 15	Peabody-Spelling--TB
72	PEA 19	Peabody-General Information--TS
73	PEA 21	Peabody-General Information--TB
74	SORT 1	Slosson Oral Reading Test--TS
75	SORT 2	Slosson Oral Reading Test--EC
76	SORT 3	Slosson Oral Reading Test--TB
77	SORT 4	Slosson Oral Reading Test--AR
78	SORT 5	Slosson Oral Reading Test--TD
79	WISC 1	Weschler-Full Scale--TS
80	WISC 2	Weschler-Full Scale--EC
81	WISC 3	Weschler-Full Scale--TB
82	WISC 7	Weschler-Verbal Scale--TS
83	WISC 8	Weschler-Verbal Scale--EC
84	WISC 11	Weschler-Verbal Scale--TD
85	WISC 13	Weschler-Performance Scale--TS
86	WISC 14	Weschler-Performance Scale--EC
87	WISC 17	Weschler-Performance Scale--TD

Key: TS = Test Scores
EC = Examiner's Comments
TB = Test Booklet
AR = Audio Recording
TD = Test Directions

APPENDIX G

PROCEDURES FOR TWO STUDIES: OS '77 and OS '78.3

Table G1.--Procedures for two studies of diagnostic decision making using simulated cases and eight clinicians.

OS '77	OS '78.3
1. Instructions and practice with sample simulated case.	1. Instructions and practice with sample simulated case.
2. Initial contact information given on simulated case.	2. Initial contact information given on simulated case.
3. Observational session directed by experimenter and recorded by a clinical observer and tape recording. Subjects collected data from "boxed" simulated case using inventory of available data. Subjects <u>were</u> encouraged to verbalize thinking.	3. Observational session directed and recorded by experimenter only and tape recording. Subjects collected data from "boxed" simulated case with <u>no</u> inventory of available data. Subjects were <u>not</u> instructed to verbalize thinking.
4. Subjects prepared written diagnostic and remedial report.	4. Subjects prepared written diagnostic and remedial report.
5. There was a debriefing session, directed by a clinical observer and aided by an experimenter, in which the subjects underwent stimulated recall. Each data item that had been requested was presented along with controlled interview questions: "Why did you request this? What did it tell you?"	5. Subjects were asked to transfer written diagnosis to standardized Reading Diagnostic Check List.
6. Written diagnosis transferred to a diagnostic check list by independent judge agreement.	6. There was a debriefing session, in which the subjects underwent stimulated recall. Each data item that had been requested was listed along with written check list: "(1) Did you ask for this information because ...you had a hunch? ...you just wanted this information? ...you usually get this information? ...other?" and "(2) Did this information ...confirm a hunch? ...disconfirm a hunch? ...suggest a hunch? ...other?"

APPENDIX H

TABLES

Table H1.--Proportion of agreement on cues, Case 2.

Cue Number	Proportion Value	Cue Name	
7	.83	Bkg. 14	School Record (EC)
8	.50	Bkg. 20	Pupil Evaluation Reports (EC)
9	.50	Bkg. 26	Standardized Achievement Test Scores (EC)
16	.83	DUR 3	Durrell Oral Reading (TB)
17	.83	DUR 4	Durrell Oral Reading (AR)
21	.67	DUR 9	Durrell Silent Reading (TB)
24	.50	DUR 14	Durrell Listening Comprehension (EC)
25	.83	DUR 15	Durrell Listening Comprehension (TB)
29	.83	DUR 21	Durrell Word Recognition and Word Analysis (TB)
37	.67	DUR 33	Durrell Hearing Sounds in Words (TB)
40	.67	GMG 1	Gates-MacGinitie Vocabulary (TS)
43	.67	GMG 7	Gates-MacGinitie Comprehension (TS)
51	.67	IRA 3	Individual Reading Analysis Common Suffixes (TB)
54	.83	IRA 9	Individual Reading Analysis Consonant Blends (TB)
57	.67	IRA 15	Individual Reading Analysis Syllabication (TB)
60	.83	SORT 3	Slosson Oral Reading (TB)
64	.67	SPE 3	Spelling (TB)
70	.50	WEP 3	Wepman Auditory (TB)
72	.50	WISC 1	Weschler Full Scale (TS)
75	.83	WISC 7	Weschler Verbal (TS)
78	.67	WISC 13	Weschler Performance (TS)
88	.50	VIS 3	Visual Acuity (TB)
89	.50	AUD 3	Auditory Acuity (TB)
N = 23			

Key: TS = Test Scores; EC = Examiner's Comments; TB = Test Booklet; AR = Audio Recording; TD = Test Directions

Table H2.--Proportion of agreement on cues, Case 3.

Cue Number	Proportion Value	Cue Name	
5	.83	BKG 2	Background Parent Form (EC)
6	.83	BKG 8	Background Teacher Form (EC)
7	.50	BKG 14	Background School Record (EC)
8	.50	BKG 20	Background Pupil Progress (EC)
9	.50	BKG 25	Standardized Achievement Test Scores (TS)
10	.50	BKG 27	Standardized Achievement Test Scores (TB)
13	1.00	DUE 3	Durrell Oral Reading (TB)
14	1.00	DUR 4	Durrell Oral Reading (AR)
17	.50	DUR 8	Durrell Silent Reading (EC)
18	.67	DUR 9	Durrell Silent Reading (TB)
22	.83	DUR 15	Durrell Oral Comprehension (TB)
30	.67	DUR 27	Durrell Visual Memory of Words Intermediate (TB)
34	.83	DUR 33	Durrell Spelling (TB)
41	.83	GMG 1	Gates-MacGinitie Vocabulary (TS)
45	.83	GMG 7	Gates-MacGinitie Comprehension (TS)
47	.67	GMG 9	Gates-MacGinitie Comprehension (TB)
55	.50	GMK 3	Gates-McKillup Blending Common Word Parts (TB)
59	.50	GMK 9	Gates-McKillup Nonsense Words (TB)
63	.60	GMK 15	Gates-McKillup Nonsense Words (TB)
66	.50	HAN 3	Handwriting (TB)
88	.83	SORT 3	Slosson Graded Word List (TB)
98	.50	UNF 3	Unfinished Sentences (TB)
108	.67	WISR 7	Weschler Verbal (TS)
110	.67	WISR 13	Weschler Performance (TS)
114	.50	AUD 2	Auditory Acuity (EC)
117	.50	VIS 3	Visual Acuity (TB)
N = 26			

Key: TS = Test Scores; EC = Examiner's Comments; TB = Test Booklet; AR = Audio Recording; TD = Test Directions

Table H3.--Proportion of agreement on cues, Case 4.

Cue Number	Proportion Value	Cue Name	
1	.50	BKG 3	Background Information Parent Form (TB)
2	.50	BKG 9	Background Information Teacher Form (TB)
3	.50	BKG 15	Background Information School Record (TB)
14	.83	DOL 3	Dolch Vocabulary Word List (TB)
18	1.00	DUR 3	Durrell Oral Reading (TB)
19	.50	DUR 4	Durrell Oral Reading (AR)
23	1.00	DUR 9	Durrell Silent Reading (TB)
32	1.00	DUR 21	Durrell Word Recognition and Word Analysis (TB)
37	.50	DUR 27	Durrell Hearing Sounds in Words Primary (TB)
41	.83	DUR 33	Durrell Visual Memory of Words Primary (TB)
45	.50	DUR 39	Durrell Intermediate Spelling List 1 (TB)
60	.67	GMK 3	Gates-McKillup Recognizing and Blending Common Word Parts (TB)
65	.50	GMK 9	Gates-McKillup Auditory Blending (TB)
70	.50	GMK 15	Gates-McKillup Giving Letter Sounds (TB)
86	.67	SORT 3	Slosson Oral Reading Graded Word List (TB)
92	1.00	WISR 1	Weschler Verbal (TS)
94	1.00	WISR 7	Weschler Performance (TS)
96	.83	WISR 13	Weschler Full Scale (TS)
97	.50	WISR 14	Weschler Full Scale (EC)
102	.50	VIS 3	Vision Test (TB)
N = 20			

Key: TS = Test Scores; EC = Examiner's Comments; TB = Test Booklet; AR = Audio Recording; TD = Test Directions

Table H4.--Percentages in total proportion of agreement (P.A.) on _cues_ for four cases.

Case Number	Total Cues Available	% & (No.) of Cues _Not_ Collected (P.A.=.0)	% & (No.) of Cues Collected by 1 Clinician (P.A.=.17)	% & (No.) of Cues Collected by 2 Clinicians (P.A.=.33)	% & (No.) of Cues Collected by 3 Clinicians (P.A.=.50)	% & (No.) of Cues Collected by 4 Clinicians (P.A.=.67)	% & (No.) of Cues Collected by 5 Clinicians (P.A.=.83)	% & (No.) of Cues Collected by 6 Clinicians (P.A.=1.0)
1 (n=6)	87	52% (45)	18% (16)	13% (11)	7% (6)	6% (5)	3% (3)	1% (1)
2 (_n_=6)	90	42% (38)	20% (18)	10% (9)	8% (7)	10% (9)	9% (8)	1% (1)
3 (_n_=6)	118	47% (56)	22% (26)	7% (8)	10% (12)	5% (6)	6% (7)	3% (3)
4 (_n_=6)	103	51% (52)	13% (13)	16% (16)	10% (10)	3% (3)	3% (3)	6% (6)

Table H5.--Percentages in total proportion of agreement (P.A.) on diagnostic statements for four cases.

Case Number	Total Statements on Check List	% & (No.) of Statements Not Mentioned (P.A.=.0)	% & (No.) of Statements Mentioned by 1 Clinician (P.A.=.17)	% & (No.) of Statements Mentioned by 2 Clinicians (P.A.=.33)	% & (No.) of Statements Mentioned by 3 Clinicians (P.A.=.50)	% & (No.) of Statements Mentioned by 4 Clinicians (P.A.=.67)	% & (No.) of Statements Mentioned by 5 Clinicians (P.A.=.83)	% & (No.) of Statements Mentioned by 6 Clinicians (P.A.=1.0)
1 (n=6)	507	81% (411)	14% (72)	3% (17)	1% (6)	0% (0)	0% (1)	0% (0)
2 (n=6)	507	82% (417)	12% (63)	4% (18)	1% (7)	0% (2)	0% (0)	0% (0)
3 (n=6)	507	77% (388)	18% (91)	5% (23)	1% (4)	0% (1)	0% (0)	0% (0)
4 (n=6)	507	77% (390)	16% (82)	5% (25)	1% (5)	0% (2)	0% (3)	0% (0)

Table H6.--Commonality of agreement on <u>cues</u> collected in six diagnoses on each of four simulated reading cases.

Commonality	Frequency				Row Totals
	Case 1	Case 2	Case 3	Case 4	
.10-.20	0	0	0	0	0
.21-.30	0	1	0	0	1
.31-.40	0	0	1	0	1
.41-.50	0	0	0	0	0
.51-.60	1	0	0	1	2
.61-.70	1	2	0	1	4
.71-.80	4	0	3	1	8
.81-.90	0	2	1	3	6
.91-1.0	0	1	1	0	2
n =	6	6	6	6	
M =	.71	.74	.77	.77	
Range =	.48/.82	.30/.93	.39/.95	.62/.86	
SD =	.13	.24	.20	.11	

Mean total = .75

SD = .03

Table H7.--Commonality of agreement on diagnostic statements made in six diagnoses on each of four simulated reading cases.

Commonality	Frequency				Row Totals
	Case 1	Case 2	Case 3	Case 4	
.10-.20	0	0	0	0	0
.21-.30	0	1	1	0	2
.31-.40	1	0	2	1	4
.41-.50	2	1	0	0	3
.51-.60	1	2	1	1	5
.61-.70	0	2	2	3	7
.71-.80	2	0	0	1	3
Above .80	0	0	0	0	0
n =	6	6	6	6	
M =	.55	.54	.47	.62	
Range =	.37/.76	.25/.66	.28/.64	.42/.78	
SD =	.16	.15	.15	.12	

Mean total = .55

SD = .05

Table H8.--Inter/intraclinician agreement (Phi and Porter) on <u>cues</u> collected on four simulated reading cases.

	Frequency				Row Totals
	Case 1	Case 2	Case 3	Case 4	
	Phi Coefficient				
Negative values	1	1	0	0	2
.00-.20	4	4	2	1	11
.21-.30	4	1	3	1	9
.31-.40	6	1	4	7	18
.41-.50	0	4	4	4	12
.51-.60	0	3	1	2	6
Above .60	0	1	1	0	2
_n_ =	15	15	15	15	
_M_ =	.30	.35	.38	.42	
Range =	-.04/.55	-.18/.87	.04/.63	.21/.60	
_SD_ =	.18	.28	.15	.11	
					Mean total = .36
					_SD_ = .05
	Porter Index				
Negative values	0	0	0	0	0
.00-.20	5	5	3	0	13
.21-.30	5	0	5	4	14
.31-.40	2	4	5	8	19
.41-.50	3	5	1	3	12
.51-.60	0	0	1	0	1
Above .60	0	1	0	0	1
_n_ =	15	15	15	15	
_M_ =	.28	.36	.33	.37	
Range =	.09/.46	.05/.83	.09/.54	.23/.52	
_SD_ =	.12	.20	.12	.08	
					Mean total = .34
					_SD_ = .04

Table H9.--Inter/intraclinician agreement (Phi and Porter) on diagnostic statements made on four simulated reading cases.

	Frequency				Row Totals
	Case 1	Case 2	Case 3	Case 4	
		Phi Coefficient			
Negative values	0	1	2	0	3 3
.00-.20	13	13	12	13	50 50
.21-.30	2	0	1	2	5 5
.31-.40	0	0	0	0	0 0
.41-.50	0	1	0	0	1 1
Above .50	0	0	0	0	0 0
n =	15	15	15	15	
M	.10	.12	.07	.16	
Range	.00/.28	-.03/.43	-.05/.26	.04/.31	
SD =	.08	.11	.08	.07	
					Mean total = .11
					SD = .04
		Porter Index			
Negative values	0	0	0	0	0
.00-.20	15	15	15	15	60
Above .20	0	0	0	0	0
n =	15	15	15	15	
M =	.07	.09	.06	.11	
Range =	.02/.17	.00/.29	.00/.16	.05/.20	
SD =	.05	.07	.04	.05	
					Mean total = .08
					SD = .02

Table H10.--Correlation coefficient for cue times, including missing data, for Case 1.

Subject Number	Run Number	Value of r
.03	1	.427
103	3	-.016
107	1	.415
107	3	.232
108	2	.401
102	2	.364
		Mean total = .30
		SD = .17

Table H11.--Correlation coefficient for cue times, including missing data, for Case 2.

Subject Number	Run Number	Value of r
102	1	.390
102	3	.374
108	1	.461
108	3	.401
104	2	.260
106	2	.051
		Mean total = .32
		SD = .15

Table H12.--Correlation coefficient for cue times, including missing data, for Case 3.

Subject Number	Run Number	Value of r
101	1	.234
101	3	.715
105	1	.287
105	3	.404
103	2	.263
107	2	.266
		Mean total = .36
		SD = .18

Table H13.--Correlation coefficient for cue times, including missing data, for Case 4.

Subject Number	Run Number	Value of r
104	1	.591
104	3	.441
106	1	.470
106	3	.223
101	2	.560
105	2	.439
		Mean total = .45
		SD = .13

Table H14.--Correlation coefficient for diagnostic statement times, including missing data, for Case 1.

Subject Number	Run Number	Value of *r*
103	1	.200
103	3	.144
107	1	.190
107	3	.317
108	2	.231
102	2	.132
		Mean total = .20
		SD = .07

Table H15.--Correlation coefficient for diagnostic statement times, including missing data, for Case 2.

Subject Number	Run Number	Value of *r*
102	1	.198
102	3	.227
108	1	.317
108	3	.292
104	2	.142
106	2	.303
		Mean total = .25
		SD = .07

Table H16.--Correlation coefficient for diagnostic statement times, including missing data, for Case 3.

Subject Number	Run Number	Value of *r*
101	1	.133
101	3	.156
105	1	.213
105	3	.322
103	2	.156
107	2	.320
		Mean total = .22
		SD = .09

Table H17.--Correlation coefficient for diagnostic statement times, including missing data, for Case 4.

Subject Number	Run Number	Value of *r*
104	1	.199
104	3	.248
106	1	.296
106	3	.283
101	2	.339
105	2	.404
		Mean total = .30
		SD = .07

Table H18.--Average of percentage of number of times each relationship appeared for cues--Case 2.

Rel. Number	Relationship	Subject 102 Run 1	Subject 102 Jun 3	Subject 108 Run 1	Subject 108 Run 3	Subject 104 Run 2	Subject 106 Run 2	Row Total	M	SD
1	Hunch	13.58	5.56	4.17	14.81	2.27	3.70	44.09	7.35	5.42
2	Just Wanted Info.	3.70	22.22	29.17	23.70	15.15	0.0	93.94	15.66	11.65
3	Usually Get Info.	9.26	6.67	6.25	7.41	27.27	21.30	78.16	13.03	9.98
4	Other	0.0	0.0	6.25	0.0	0.0	1.39	7.64	1.27	2.50
5	Confirm Hunch	9.88	11.67	14.58	14.81	25.00	18.06	94.00	15.67	5.37
6	Disconfirm Hunch	6.79	2.22	0.0	3.70	0.0	0.0	12.71	2.12	2.75
7	Suggest Hunch	9.26	9.44	18.75	7.41	21.21	21.30	87.37	14.56	6.52
8	Other	1.85	0.0	0.0	9.63	0.0	0.0	11.48	1.91	3.85
9(1+5)	Hunch & Confirmed	8.64	22.22	12.50	18.52	0.0	9.26	71.14	11.86	7.87
10(1+6)	Hunch & Disconfirmed	3.70	0.0	0.0	0.0	0.0	0.0	3.70	.62	1.51
11(2+5)	Wanted Info. & Confirmed	0.0	20.00	0.0	0.0	4.55	0.0	24.55	4.09	8.00
12(2+7)	Info. & Suggested Hunch	0.0	0.0	0.0	0.0	0.0	2.78	2.78	.46	1.14
13(3+5)	Usually Get & Confirmed Hunch	3.70	0.0	0.0	0.0	2.27	11.11	17.08	2.85	4.33
14(3+7)	Usually Get & Suggested Hunch	3.70	0.0	0.0	0.0	2.27	11.11	17.08	2.85	4.33

Table H19.--Average of percentage of number of times each relationship appeared for cues--Case 3.

Rel. Number	Relationship	Subject 101 Run 1	Subject 101 Run 3	Subject 105 Run 1	Subject 105 Run 3	Subject 103 Run 2	Subject 107 Run 2	Row Total	M	SD
1	Hunch	5.56	8.26	16.95	3.95	15.38	0.0	50.10	8.35	6.64
2	Just Wanted Info.	0.0	8.70	3.45	14.04	3.85	3.97	34.01	5.67	4.95
3	Usually Get Info.	0.0	0.0	12.07	11.40	11.54	22.62	57.63	9.61	3.57
4	Other	0.0	3.62	3.45	0.0	0.0	0.0	7.07	1.18	1.83
5	Confirm Hunch	8.33	33.04	9.20	19.30	9.62	31.35	110.84	18.47	11.37
6	Disconfirm Hunch	0.0	0.0	9.48	0.0	5.77	9.52	24.77	4.13	4.72
7	Suggest Hunch	33.33	9.42	3.45	2.63	23.08	4.76	76.67	12.78	12.60
8	Other	0.0	4.35	3.45	0.0	0.0	8.73	16.53	2.76	3.51
9(1+5)	Hunch & Confirmed	44.44	26.81	17.82	31.58	3.85	4.76	129.26	21.54	15.83
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	0.0	0.0	0.0	7.89	0.0	4.76	12.65	2.11	3.41
12(2+7)	Info. & Suggested Hunch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13(3+5)	Usually Get & Confirmed Hunch	0.0	0.0	1.72	1.32	0.0	0.0	10.04	1.67	2.72
14(3+7)	Usually Get & Suggested Hunch	0.0	1.45	1.72	0.0	19.23	0.0	22.40	3.73	7.63

Table H20.--Average of percentage of number of times each relationship appeared for cues--Case 4.

Rel. Number	Relationship	Subject 104 Run 1	Subject 104 Run 3	Subject 106 Run 1	Subject 106 Run 3	Subject 101 Run 2	Subject 105 Run 2	Row Total	M	SD
1	Hunch	20.83	16.67	3.85	11.90	25.90	4.55	83.70	13.95	8.86
2	Just Wanted Info.	5.00	9.22	8.97	3.17	0.0	12.12	38.50	6.42	4.49
3	Usually Get Info.	10.83	18.04	21.86	22.38	1.28	3.79	78.18	13.03	9.15
4	Other	0.0	0.0	0.0	0.0	0.0	1.52	1.52	.25	.62
5	Confirm Hunch	25.83	30.98	19.23	20.40	20.00	18.18	134.62	22.44	4.96
6	Disconfirm Hunch	0.0	0.0	4.49	7.14	14.10	0.0	25.73	4.29	5.65
7	Suggest Hunch	2.50	19.22	8.91	8.81	14.36	9.60	63.40	10.57	5.68
8	Other	0.0	0.0	0.0	7.14	7.69	0.0	14.83	2.48	3.83
9(1+5)	Hunch & Confirmed	5.00	2.94	7.69	9.52	16.67	37.12	78.94	13.16	12.65
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	3.75	0.0	15.38	0.0	0.0	6.06	25.19	4.20	6.03
12(2+7)	Info. & Suggested Hunch	0.0	2.94	0.0	0.0	0.0	3.03	5.97	1.0	1.54
13(3+5)	Usually Get & Confirmed Hunch	2.92	0.0	1.92	0.0	0.0	0.0	4.84	.81	1.29
14(3+7)	Usually Get & Suggested Hunch	3.33	0.0	3.85	0.0	0.0	1.01	8.19	1.37	1.78

Table H21.--Average of percentage of number of times each relationship appeared for diagnostic statements--Case 2.

Rel. Number	Relationship	Subject 102 Run 1	Subject 102 Run 3	Subject 108 Run 1	Subject 106 Run 3	Subject 104 Run 2	Subject 106 Run 2	Row Total	M	SD
1	Hunch	28.85	10.53	5.36	11.71	3.70	6.82	66.97	11.16	9.18
2	Just Wanted Info.	4.81	11.53	35.71	19.82	15.12	0.0	87.04	14.51	12.58
3	Usually Get Info.	5.77	10.53	10.71	6.31	29.14	24.24	86.70	14.45	9.82
4	Other	0.0	0.0	5.36	0.0	0.0	2.27	7.63	1.27	2.20
5	Confirm Hunch	20.19	16.84	17.86	17.57	16.30	18.18	106.94	17.82	1.35
6	Disconfirm Hunch	10.58	5.26	0.0	2.70	0.0	0.0	18.54	3.09	4.23
7	Suggest Hunch	9.62	14.21	20.54	4.73	30.56	27.27	106.93	17.82	10.11
8	Other	3.85	0.0	0.0	21.62	0.0	0.0	25.47	4.25	8.65
9(1+5)	Hunch & Confirmed	7.69	19.47	4.46	15.54	0.0	4.55	51.71	8.62	7.41
10(1+6)	Hunch & Disconfirmed	3.85	0.0	0.0	0.0	0.0	0.0	3.85	.64	1.57
11(2+5)	Wanted Info. & Confirmed	0.0	11.58	0.0	0.0	0.74	0.0	12.32	2.05	4.68
12(2+7)	Info. & Suggested Hunch	0.0	0.0	0.0	0.0	0.0	4.55	4.55	.76	1.86
13(3+5)	Usually Get & Confirmed Hunch	3.85	0.0	0.0	0.0	3.70	9.09	16.64	2.77	3.61
14(3+7)	Usually Get & Suggested Hunch	0.96	0.0	0.0	0.0	.74	3.03	4.73	.73	1.13

Table H22.--Average of percentage of number of times each relationship appeared for diagnostic statements--Case 3.

Rel. Number	Relationship	Subject 101 Run 1	Subject 101 Run 3	Subject 105 Run 1	Subject 105 Run 3	Subject 103 Run 2	Subject 107 Run 2	Row Total	M	SD
1	Hunch	28.57	13.04	24.29	11.72	13.04	0.0	90.66	15.11	10.14
2	Just Wanted Info.	0.0	4.35	1.43	16.41	7.14	6.45	35.78	5.96	5.82
3	Usually Get Info.	0.0	0.0	6.00	8.59	16.07	24.19	54.85	9.14	9.51
4	Other	0.0	3.04	5.71	0.0	0.0	0.0	8.75	1.46	2.41
5	Confirm Hunch	7.14	43.04	11.43	23.44	23.21	37.10	145.36	24.23	13.98
6	Disconfirm Hunch	0.0	0.0	14.86	0.0	10.71	8.06	36.63	5.61	6.51
7	Suggest Hunch	21.43	9.57	8.57	1.56	15.09	3.23	59.45	9.91	7.43
8	Other	0.0	5.22	0.57	0.0	0.0	6.45	12.24	2.04	2.97
9(1+5)	Hunch & Confirmed	42.86	20.29	22.86	17.19	0.45	1.61	105.26	17.54	15.64
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	0.0	0.0	0.0	1.56	0.0	3.23	4.79	.80	1.34
12(2+7)	Info. & Suggested Hunch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13(3+5)	Usually Get & Confirmed Hunch	0.0	0.0	1.43	0.78	0.0	0.0	2.21	.37	.61
14(3+7)	Usually Get & Suggested Hunch	0.0	1.45	2.86	0.0	10.71	0.0	15.02	2.50	4.18

Table H23.--Average of percentage of number of times each relationship appeared for diagnostic statements--Case 4.

Rel. Number	Relationship	Subject 104 Run 1	Subject 104 Run 3	Subject 106 Run 1	Subject 106 Run 3	Subject 101 Run 2	Subject 105 Run 2	Row Total	M	SD
1	Hunch	13.04	6.91	3.03	7.81	21.15	6.25	58.19	9.70	6.48
2	Just Wanted Info.	4.35	17.07	10.61	2.60	0.0	14.06	48.69	8.12	6.81
3	Usually Get Info.	30.43	16.26	30.30	33.33	3.85	9.38	123.55	20.59	12.47
4	Other	0.0	0.0	0.0	0.0	0.0	0.78	0.78	.13	.32
5	Confirm Hunch	27.54	25.00	18.18	22.81	22.44	29.69	145.66	24.28	4.08
6	Disconfirm Hunch	0.0	0.0	6.06	6.25	9.62	0.0	21.93	3.66	4.20
7	Suggest Hunch	5.07	27.44	18.18	14.06	24.36	16.67	105.78	17.63	7.92
8	Other	0.0	0.0	0.0	6.88	1.92	0.0	8.80	1.47	2.76
9(1+5)	Hunch & Confirmed	4.35	2.44	4.04	6.25	12.82	17.19	47.09	7.85	5.84
10(1+6)	Hunch & Disconfirmed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11(2+5)	Wanted Info. & Confirmed	6.52	0.0	7.07	0.0	0.0	1.82	15.41	2.57	3.35
12(2+7)	Info. & Suggested Hunch	0.0	2.44	0.0	0.0	0.0	1.04	3.48	.58	1.00
13(3+5)	Usually Get & Confirmed Hunch	4.35	0.0	1.01	0.0	0.0	0.0	5.36	.89	1.74
14(3+7)	Usually Get & Suggested Hunch	4.35	0.0	1.52	0.0	0.0	3.13	9.00	1.50	1.87

Table H24.--Individual case data (mean and standard deviation) on Porter statistic, including number of cues and diagnostic statements in domains, for two studies.

	CUES								DIAGNOSTIC STATEMENTS							
	OS '77 Study				OS '78.3 Study				OS '77 Study				OS '78.3 Study			
CASES:	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Inter/intra (Porter)																
M	.34	.35	.30	.28	.28	.36	.33	.37	.10	.12	.11	.13	.07	.09	.06	.11
SD	.13	.12	.11	.07	.12	.20	.12	.08	.07	.06	.11	.07	.05	.07	.04	.05
Intra (Porter)																
M	.40	.50	.44	.29	.25	.63	.35	.44	.21	.20	.30	.22	.17	.22	.10	.19
SD	.23	.17	.01	.06	.16	.29	.04	.11	.06	.07	.13	.01	.00	.11	.09	.02
No. of Items in Domain Computed	50	62	73	78	87	90	118	103	48	44	41	40	507	507	507	507

APPENDIX I

OBSERVATIONAL STUDY DATA-ANALYSIS SYSTEM (OSDAS) STATISTICS

OBSERVATIONAL STUDY DATA-ANALYSIS SYSTEM (OSDAS) STATISTICS

The following statistics are all calculated by (1) a computer statistical-analysis system (Observational Study Data-Analysis System) developed and maintained by the Institute for Research on Teaching, Michigan State University (Wagner et al., 1979) and (2) an interactive statistical-analysis system (MIDAS) at the University of Michigan. A random sampling of data was computed by hand across numerous observational studies, and the system was found to be operating accurately. Some portions of the data, such as mean and standard deviation, were calculated with a Litronix 2270R hand calculator.

Proportional Agreement

Given a domain for cues or diagnostic statements (or remediations) for a given case, proportional agreement is the proportion of clinicians who mentioned each cue or diagnostic statement. One proportion is computed for each cue or diagnostic statement in the domain.

The statistic is bounded by 0 and 1 and is calculated by determining the number of clinicians who collected the same cues or made the same diagnostic statements for a given case. That number is divided by the total number of clinicians who interacted with the case. The resulting statistic indicates the proportion of clinicians who collected the same cues or made the same diagnostic statements for a specific case. The calculation is for the most frequently collected cues and most frequently made diagnostic statements. The formal computation is as follows:

$$P.A. = \frac{C_{ji}}{C_j}$$

where C_{ji} = number of clinicians mentioning the i^{th} category
C_j = total number of clinicians for a given case.

For example, if two clinicians, of a total of six diagnosing the same case, mentioned the diagnostic statement "Contextual Word Recognition--Weakness," the proportional agreement would be:

$$P.A. = \frac{\text{no. of clinicians mentioning cue (statement)}}{\text{total number of clinicians}}$$

$$P.A. = \frac{2}{6} = .33$$

Commonality

Given a domain for cues or diagnostic statements for a given case, the commonality statistic is a measure of agreement between one clinical session and all other clinical sessions for a given case; e.g., an individual is compared with a group.

The statistic is bounded by 0 and 1, and only the proportional-agreement statistic is used in calculating an individual's score. A value of *x* for a specific clinician implies that he has collected in his session, for a given case, roughly *x* percent of those cues most frequently collected by the group for that same case. If, for example, a clinician has a commonality score of .47 it means that the clinician has collected in his session, for a given case, roughly 47 percent of those cues most frequently collected by the group for the same case. The same analysis can be applied to diagnostic statements.

Interclinician Agreement

One of the indices used to describe interclinician agreement was the Phi coefficient. The Phi, denoted by ϕ, is the traditional Pearson product-moment coefficient of correlation for nominal dichotomous data with no assumptions concerning the shape of the distribution of scores. One Phi coefficient was computed for each pair of clinicians.

To compute the interclinician correlation, two requirements were met. First, a domain of statements was defined. (See Appendix C.) Second, a determination was made as to which diagnostic statements were present in a diagnosis or absent from a diagnosis. When two clinicians were compared, for convenience, the data could be tabulated in 2 x 2 contingency tables showing the joint occurrences of pairs of scores (+ 1) using frequencies. For this study, the cues (Cx) or diagnostic statements (Dx) mentioned by one clinician were compared with those mentioned by a second clinician for the same case. This comparison is illustrated in Figure I1.

The calculation of ϕ is derived from the contingency table as follows:

$$\text{Phi} = \frac{(ad - bc)}{\sqrt{(a+c)(b+d)(c+d)(a+b)}}$$

(See Figure I2.)

Clinician B, SIMCASE Y	Clinician A, SIMCASE Y: PRESENT (+)	Clinician A, SIMCASE Y: ABSENT (-)
PRESENT (+)	Frequency count of items in the domain present in both clinicians' Cx/Dx a	Frequency count of items present in clinician B's Cx/Dx but not in clinician A's Cx/Dx b
ABSENT (-)	Frequency count of items in the domain present in clinician A's Cx/Dx but not in clinician B's Cx/Dx c	Frequency count of items in the domain absent in both clinicians' Dx/Dx d

Figure I1.--Contingency table for calculation of Phi--Example 1.

Clinician B, SIMCASE Y	Clinician A, SIMCASE Y: +	Clinician A, SIMCASE Y: -	
+	a (+ +)	b (+ -)	a + b
-	c (- +)	d (- -)	c + d
	a + c	b + d	N

Figure I2.--Contingency table for calculation of Phi--Example 2.

The Phi statistic is bounded by -1 (when items are in cells b and c only) and 1 (when items are in cells a and d only), when the distributions in the marginals are equal. In all other instances, the maximum and minimum values will be less than 1 and greater than -1. An example of a completed table is shown in Figure I3.

Clinician A

Clinician B	+	-	
+	2 a	1 b	3
-	1 c	4 d	5
	3	5	8

Cues (C) of Clinician A SIMCASE Y	Cues (C) of Clinician B SIMCASE Y	Domain of Cues (C)
		C1
C1	C3	C2
		C3
C3	C5	C4
		C5
C5	C7	C6
		C7
		C8

Figure I3.--An example of a completed contingency table.

The second of the two indices used in this study to describe the interclinician agreement was $\frac{A}{A+B+C}$, which describes the proportion of agreement where the base was the total number of cues collected (or diagnostic statements made) for which one or the other or both clinicians collected the cue or made the diagnostic statement.* The upper bound of the index would be the value of the index $\frac{A+D}{A+B+C+D}$, which describes the proportion of agreement where the base was the total number of cues in the particular cue domain (or diagnostic statements in the statement domain). The Porter Index, $\frac{A}{A+B+C}$, excludes clinician agreement not to request a cue or make a diagnostic statement, i.e., the "d" cell (- -) in the 2 x 2 contingency table.

Intraclinician Agreement

The same two statistical measures, the Phi and the Porter, that were used to determine interclinician agreement were used for intraclinician agreement.

The intraclinician-agreement statistic reflects the agreement of one clinician's collection of cues for a specific case with his own collection of cues on an alternate form of the same case. It compares the presence or absence of certain cues collected by a given clinician (C) on a specific case at Time 1 (T_1) to the presence or absence of certain cues collected by that same clinician on the same case (alternate form) at a later time (T_2). The same analysis can be used to compute the intraclinician agreement on

---

*Statistic developed by Andrew Porter, Michigan State University, for the National Day Care Study (Wilcox, 1977, pp. 54-60).

diagnostic statements selected. One Phi coefficient is computed for each pair of sessions. This comparison is summarized in Figure I4.

Clinician C
SIMCASE Y, FORM 1

Clinician C SIMCASE Y, FORM 2		PRESENT (+)	ABSENT (-)
	PRESENT (+)	Frequency count of items present in the domain in both sessions for FORM 1 and FORM 2 of SIMCASE Y a	Frequency count of items in the domain present in the session for FORM 2 SIMCASE but not in FORM 1 SIMCASE Y b
	ABSENT (-)	Frequency count of items present in the session for FORM 1 SIMCASE but not in FORM 2 SIMCASE Y c	Absent in both sessions for FORM 1 and FORM 2 of SIMCASE Y d

Figure I4.--Contingency table for calculation of intra Phi--Example 1.

The calculation of the Phi is derived from the preceding contingency table in the following manner:

$$\text{Phi} = \frac{(ad - bc)}{\sqrt{(a+c)(b+d)(c+d)(a+b)}}$$

(See Figure I5.)

Clinician A SIMCASE Y, FORM 2		Clinician A SIMCASE Y, FORM 1 +	-	
	+	a (+ +)	b (+ -)	a + b
	-	c (- +)	d (- -)	c + d
		a + c	b + d	N

Figure I5.--Contingency table for calculation of intra Phi--Example 2.

The statistic is bounded by -1 (when items are in cells b and c only) and 1 (when items are in cells a and d only), when the distributions in the marginals are equal. In all other instances, the maximum and minimum values will be less than 1 and greater than -1.

An example of a completed contingency table is shown in Figure I6.

Clinician A SIMCASE Y, FORM 2 \ Clinician A SIMCASE Y, FORM 1	+	-	
+	2	2	4
-	1	3	4
	3	5	8

Cues (C) of Clinician A SIMCASE Y, FORM 1	Cues (C) of Clinician A SIMCASE Y, FORM 2	Domain of Cues (C)
		C1
C1	C1	C2
		C3
C3	C3	C4
	C6	C5
C4	C7	C6
		C7
		C8

Figure I6.--An example of a completed contingency table.

The Porter Index is also used to compute intraclinician agreement, where a clinician is compared with himself on cues collected or diagnostic statements made for alternate forms of the same case.

Directions for Computing Relationship Responses

The means and standard deviations were computed (for each run for each subject) on the hypothesis-directed inquiry responses by the following procedures, using a hand calculator:

1. Compute the sums of the individual columns for relationship numbers 1, 5, 6, 9, and 10 from each of the relationship tables. (See Table 20 and Appendix H.)

2. Compute the means (of the sums of the individual columns for relationship numbers 1, 5, 6, 9, and 10).

3. Compute the standard deviation (of the variability of the value for each run from the mean of the overall sum of relationships 1, 5, 6, 9, and 10) from the relationship table.

The means and standard deviations were computed for cue-directed inquiry on each run for each subject according to the following procedures, using a hand calculator:

1. Compute the sums of the individual columns for relationship numbers 2, 3, 4, 7, 8, 11, 12, 13, and 14 from each of the relationship tables.

2. Compute the means (of the sums of the individual columns for relationship numbers 2, 3, 4, 7, 8, 11, 12, 13, and 14) from the relationship tables.

3. Compute the standard deviation (of the variability of the value for each run from the mean of the overall sum of relationships 2, 3, 4, 7, 8, 11, 12, 13, and 14) from the relationship tables.

The same two sets of procedures were used to compute hypothesis-directed inquiry and cue-directed inquiry for relationships appearing for diagnostic statements (Table 23, based on relationship tables for diagnostic statements).

BIBLIOGRAPHY

BIBLIOGRAPHY

Barrows, H. S.; Feightner, J. W.; Neufeld, V. R.; and Norman, G. R. Data Gathering Problem Solving and Management-Observation of the Clinical Encounter in Time Segments. Hamilton, Canada: McMasters University, 1976.

Bell, A. E., and Aftanas, M. S. "Some Correlates of Reading Retardation." Perceptual and Motor Skills 35 (October 1972): 659-67.

Bond, G. L., and Tinker, M. A. Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, Inc., 1957.

Brown, V., and Botel, M. Dyslexia: Definition or Treatment? Bloomington, Ind.: ERIC/CRIER, Reading Review Series, 1972.

Burnett, R. "The Diagnostic Problem Solving Proficiency of Elementary Teachers in Teaching Reading." Ph.D. dissertation, Indiana University, 1961.

Carter, H. L. J. "Determining Consequential Factors in Diagnosis." In Reading Diagnosis and Evaluation, Proceedings of the Thirteenth Annual Convention of the International Reading Association. Edited by D. L. DeBoer. Newark, Delaware: International Reading Association, 1970.

________, and McGinnis, D. J. Diagnosis and Treatment of the Disabled Reader. London: The Macmillan Co., 1970.

Chall, J. S. "A Decade of Research on Reading and Learning Disabilities." In What Research Has to Say About Reading Instruction. Edited by S. J. Samuels. Newark, Delaware: International Reading Association, 1978.

Clarkson, G. P. E. Portfolio Selection: A Simulation of Trust Investment. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Cureton, E. E. "Note on PHI/PHI(MAX)." Psychometrica 24 (1959): 89-91.

DeBoer, J. J., and Dallmann, M. The Teaching of Reading. New York: Holt, Rinehart and Winston, Inc., 1960.

DeDombal, F. T.; Horrocks, J. C.; Clamp, S. E.; and Storr, J. E. "Simulation Techniques and Computer-Aided Teaching of the Clinical Diagnostic Process: Five Years Experience." Medinfo 74 Proceedings, pp. 247-52. Amsterdam: North Holland Publishing Co., 1974.

DeDombal, F. T.; Leaper, D. J.; Staniland, J. R.; McCann, A. P.; and Horrocks, J. C. "Computer-Aided Diagnosis of Acute Abdominal Pain." British Medical Journal 2 (1972): 9-13.

DeGroot, A. D. "Perception and Memory Versus Thought: Some Old Ideas and Recent Findings. In Problem Solving: Research, Method, and Theory, pp. 19-20. Edited by B. Kleinmuntz. New York: John Wiley and Sons, 1966.

________. Thought and Choice in Chess. The Hague: Mouton, 1965.

Downey, N. M., and Heath, R. W. Basic Statistical Methods. New York: Harper and Row, 1965.

Durrell, D. D. The Improvement of Basic Reading Abilities. Yonkers, N.Y.: World Book Co., 1940.

Ekwall, E. Diagnosis and Remediation of the Disabled Reader. Boston: Allyn and Bacon, Inc., 1976.

Elstein, A. S. "Clinical Judgment: Psychological Research and Medical Practice." Science 194 (November 1976): 696-700.

________. "Decision Analysis and the Psychology of Decision Making." East Lansing: Office of Medical Education Research and Development, Michigan State University, 1977.

________. "A Process Tracing Analysis of Medical Problem Solving." Paper presented to the Annual Meeting of the American Psychological Association, Chicago, Illinois, September 3, 1975.

________; Kagan, N.; Shulman, L. S.; Jason, H.; and Loupe, M. J. "Methods and Theory in the Study of Medical Inquiry." Journal of Medical Education 47 (1972): 85-92.

Elstein, A. S.; Shulman, L. S.; Sprafka, S. H.; Jason, H.: Kagan, N.; Allal, L. K.; Gordon, M. J.; Loupe, M. J.; and Jordan, R. D. "An Analysis of the Medical Inquiry Process: Final Report to the Bureau of Health Manpower, OMERAD." East Lansing: College of Human Osteopathic Medicine, Michigan State University, 1976.

________. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge: Harvard University Press, 1978.

Elstein, A.; Shulman, L.; Vinsonhaler, J.; Wagner, C.; and Bader, L. "Characteristics of the Clinical Problem Solving Model and Its Relevance to Education" (Res. Ser. No. 27). East Lansing: Institute for Research on Teaching, Michigan State University, 1977.

Feinstein, A. R. Clinical Judgment. Baltimore: Williams and Wilkins, 1967.

Gil, C.; Hoffmeyer, E. M.; VanRoekel, J. L.; and Weinshank, A. B. "Clinical Problem Solving in Reading: Theory and Research" (Res. Ser. No. 45). East Lansing: Institute for Research on Teaching, Michigan State University, 1979.

Glaser, R.; Damrin, D. E.; and Gardner, F. M. "The Tab Item: A Technique for the Measurement of Proficiency in Diagnostic Problem Solving Tasks." Educational and Psychological Measurement 14 (1954): 283-93.

Glass, G. V., and Stanley, J. C. Statistical Methods in Education. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

Gomez, M. R. "Specific Learning Disorders in Childhood." Psychiatric Annals 2 (May 1972): 49-65.

Harris, A. J. "Diagnosis of Reading Disabilities." In Corrective Reading in Classroom and Clinic, pp. 80-87. Supplementary Educational Monographs, No. 79. Edited by H. M. Robinson. Chicago: University of Chicago Press, December 1953; reprint ed., New York: Harper and Row, Publishers, 1972.

________. How to Increase Reading Ability. 4th ed. New York: Longmans, Green, and Co., 1961.

________, and Roswell, F. G. "Clinical Diagnosis of Reading Disability." Journal of Psychology 36 (1953): 323-40.

Harris, A. J., and Sipay, E. R. How to Increase Reading Ability: A Guide to Developmental and Remedial Methods. 6th ed., rev. and enl. New York: David McKay Co., 1975.

Hemphill, J. K.; Griffiths, D. E.; and Frederiksen, N. Administrative Performance and Personality. New York: Bureau of Publications, Teachers College, Columbia University, 1962.

Hoffmeyer, E. M. "A Study of the Process and Outcomes of Diagnostic Problem Solving Among Reading Clinicians." Unpublished proposal for the degree of Ph.D., Michigan State University, October 1978.

________, and Bader, L. A. "A Study of the Diagnostic Decision-Making and the Problem-Solving Behavior of Reading Clinicians as They Diagnose Reading Disability Cases." Unpublished paper, Michigan State University, May 1978.

Jan-Tausch, E. "Reading Diagnosis--The Essential Ingredient in Teaching Every Child to Read." In Diagnostic Viewpoints in Reading, pp. 1-7. Edited by R. Leibert. Newark, Del.: International Reading Association, 1971.

Kitchen, F.; Polin, R.; and Salters, D. Data Processor's Manual for the Clinical Studies Project of the Institute for Research on Teaching. East Lansing: Michigan State University, February 1979; rev. ed. June 1979.

Kleinmuntz, B. "The Processing of Clinical Information by Man and Machine." In Formal Representation of Human Judgment. Edited by B. Kleinmuntz. New York: Wiley, 1968.

Lanier, J. E., and Shulman, L. S. Technical Proposal submitted to and approved for funding by NIE. East Lansing: Institute for Research on Teaching, College of Education, Michigan State University, April 1976.

Lee, A., and Weinshank, A. "CLIPIR Pilot Observational Study of Reading Diagnosticians" (Res. Ser. No. 14). East Lansing: Institute for Research on Teaching, Michigan State University, 1976.

Norman, G.; Barrows, H. S.; Feightner, J. W.; and Neufeld, V. R. "Measuring the Outcome of Clinical Problem Solving." Unpublished paper, McMaster University, Hamilton, Ontario, 1977.

Penn, Julia M. "Reading Disability: A Neurological Deficit?" Exceptional Children 33 (December 1966): 243-48.

Rimoldi, H. J. A. "A Technique for the Study of Problem Solving." Educational and Psychological Measurement 15 (1955): 450-61.

________. "The Test of Diagnostic Skills." Journal of Medical Education 36 (1961): 73-79.

________; Devane, J.; and Haley, J. "Characterization of Processes." Educational and Psychological Measurement 21 (1961): 389-92.

Robinson, H. M. "Clinical Procedures in Diagnosing Seriously Retarded Readers." In Better Readers for Our Times, pp. 152-56. Conference Proceedings of the International Reading Association. New York: Scholastic Magazines, 1956.

Rosenthal, J. H. "Neurophysiology of Minimal Cerebral Dysfunctions." Academic Therapy 8 (Spring 1973): 291-94.

Schwartz, W. B., et al. "Decision Analysis and Clinical Judgment." American Journal of Medicine 55 (1973): 459-72.

Sherman, G. B. "Training Reading Specialists in Diagnosis" (Res. Ser. No. 31). East Lansing: Institute for Research on Teaching, Michigan State University, 1979.

Sherman, G.; Vinsonhaler, J.; and Weinshank, A. "A Model of Reading and Learning to Read." Personal correspondence to Clinical Studies Project Staff of the Institute for Research on Teaching, Michigan State University, 1978.

Shulman, L. S. "Seeking Styles and Individual Differences in Patterns of Inquiry." School Review 73 (1965): 258-66.

________, and Elstein, A. "Studies of Problem Solving, Judgement and Decision Making: Implications for Educational Research." In Review of Research in Education, pp. 3-42. Edited by F. Kerlinger. New York: Peacock, 1975.

Shulman, L. S.; Loupe, M. J.; and Piper, R. M. Studies of the Inquiry Process: Inquiry Patterns of Students in Teacher-Training Programs. Final report, Project No. 5-0597, Office of Education. East Lansing: Michigan State University, 1968.

Smith, H. P., and Dechant, E. V. Psychology in Teaching Reading. Englewood Cliffs, N.J.: Prentice-Hall, 1961.

Spache, G. D. Diagnosing and Correcting Reading Disabilities. Boston: Allyn and Bacon, Inc., 1976.

________. Good Reading for Poor Readers. 5th rev. ed. Champaign, Ill.: Gerrard Press, 1966.

Sprafka, S. A., and Elstein, A. S. "What Do Physicians Do? An Analysis of Diagnostic Reasoning." Symposium presentation of the Conference on Research in Medical Judgment, Office of Medical Education Research and Development, Michigan State University, East Lansing, Michigan, May 1974.

Stephens, E. C. A. "A Study of the Problem-Solving Behavior of Teachers as They Diagnose a Child's Reading Performance and the Effects of Experience and Training on That Behavior." Ph.D. dissertation, Michigan State University, 1978.

Taylor, W. C.; Skakun, E. N.; and Wilson, D. R. "Computerized Medical Self-Assessment." Medinfo 77 Proceedings. Edited by D. B. Shires and H. Wolf. Amsterdam: North Holland Publishing Co., 1977.

VanRoekel, J. L. "The Problem Solving Behavior of Teachers Trained in Reading and Teachers Trained in Learning Disabilities When Diagnosing Simulated Cases of Educational Difficulty in a Clinical Setting." Ph.D. dissertation, Michigan State University, in progress.

Vinsonhaler, J. F. "Clinical Problem Solving Among Reading Specialists: The Problem of Agreement on a Diagnosis." East Lansing: Institute for Research on Teaching, Michigan State University, April 1979a.

________. "The Consistency of Reading Diagnosis." (Res. Ser. No. 28). East Lansing: Institute for Research on Teaching, Michigan State University, June 1979b.

________. "Implications of an Interdisciplinary Theory of Clinical Inquiry for Diagnostic Efficacy." In Proceedings of AIMS Conference. Edited by F. Gremy and F. T. DeDombal. Amsterdam: North Holland Publishing Co., 1979c.

________. "Theory." Personal Communication paper. East Lansing: Institute for Research on Teaching, Michigan State University, 1978.

Vinsonhaler, J. F.; Wagner, C. C.; and Elstein, A. S. "The Inquiry Theory: An Information Processing Approach to Clinical Problem Solving Research and Application" (Res. Ser. No. 1). East Lansing: Institute for Research on Teaching, Michigan State University, 1977.

Vinsonhaler, J. F.; Wagner, C. C.; Elstein, A. S.; and Shulman, L. S. "The Inquiry Theoretic Approach to Clinical Problem-Solving Research and Application: Basic Management Information System (BMIS) Technical Manual," Vol. I (Res. Ser. No. 15). East Lansing: Institute for Research on Teaching, Michigan State University, 1977.

Wagner, C. C.; Kitchen, F.; Polin, R.; Vinsonhaler, J. F.; VanRoekel, J. L.; Gil, D.; and Hoffmeyer, E. M. "Observational Studies Data Analysis System (OSDAS)." IRT Technical Report. East Lansing: The Institute for Research on Teaching, Michigan State University, 1979.

White, J., and White, M. "Perceptual and Psycholinguistic Factors in Retarded and Advanced Levels of Reading and Spelling Skills." Slow Learning Child: The Australian Journal on the Education of Backward Children 19 (July 1972): 117-23.

Whitecraft, C. J. "Approaching Cooperative Research in Learning Disabilities Through Psycholinguistics." Journal of Learning Disabilities 4 (December 1971): 568-71.

Wilcox, M.; Alford, S.; Carew, J.; Gillis, G.; Porter, A.; and Stallings, J. National Day Care Home Study, Phase 1 Pilot Study Report. Menlo Park, Calif.: Stanford Research Institute, June 1977.

Wilson, R. M. Diagnostic and Remedial Reading for Classroom and Clinic. Columbus, Ohio: Charles E. Merrill Publishing Co., 1977.