THESIS

This is to certify that the

thesis entitled

The Construction and Use of Papercases to Observe the Diagnostic Problem Solving Behavior of Reading Clinicians

presented by

C. Jay Colello Stratoudakis

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Education

Byron H. Van Roekel
Major professor

Date Feb. 21, 1980

O-7639

THE CONSTRUCTION AND USE OF PAPERCASES TO OBSERVE THE DIAGNOSTIC PROBLEM SOLVING BEHAVIOR OF READING CLINICIANS

By

C. Jay Colello Stratoudakis

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

1979

ABSTRACT

THE CONSTRUCTION AND USE OF PAPERCASES TO OBSERVE THE DIAGNOSTIC PROBLEM SOLVING BEHAVIOR OF READING CLINICIANS

By

C. Jay Colello Stratoudakis

Rationale

Until quite recently, research studies on the process of diagnosing reading disabilities have been nonexistent (Spache, 1968). Since its founding in 1976, the Clinical Studies Project within the Institute for Research on Teaching at Michigan State University, has addressed the question: "How do experienced reading clinicians approach the diagnosis of children with reading problems?" This study was an extension of the research initiated by the Clinical Studies Project. It sought additional information regarding how experienced reading clinicians think about reading problems and pursued the answer to another question: "What is the most efficient and effective instrument to use for descriptive observation and eventual training of reading specialists in the process of diagnosing reading disabilities?

C. Jay Colello Stratoudakis

Research to date on the diagnostic problem-solving behavior of reading specialists has been conducted exclusively utilizing simulated cases called SIMCASES which were developed from diagnostic records of actual children with reading problems. The SIMCASE observational instrument allowed for direct observation of the reading specialist's behavior but required individual administration. No observational studies have been conducted utilizing a simulated case of reading disability in a form which allowed for group administration and observation.

Purpose

The purpose of this study was to construct and to test the usefulness of a group-administered simulated case of reading disability termed a PAPERCASE as an alternative to the SIMCASE observational instrument. Whether the PAPERCASE could be substituted for the SIMCASE in collecting data on diagnostic problem-solving performance depended upon the consistency of a clinician's performance on these two instruments.

Methodology

Twelve certified classroom teachers who earned a superior grade in a graduate-level course in reading diagnosis at Michigan State University were hired as

consultants to the Clinical Studies Project and served as subjects in this study. The twelve subjects were randomly assigned to one of three different cases of reading disability. Four subjects were assigned to Case I, Four to Case II, and four to Case III. Each subject participated in two observational sessions with a one week interval between observations. With twelve subjects observed twice, there were a total of twenty-four observational sessions divided equally among the three different cases of reading disability represented in the two observational instruments, PAPERCASE and SIMCASE.

Analysis

The diagnostic problem-solving performance of the subjects on the SIMCASES and PAPERCASES was analyzed in terms of four measures of "agreement statistics" developed by the Clinical Studies Project: Proportional Agreement, Commonality, Inter-Clinician Agreement, and Intra-Clinician Agreement (Vinsonhaler, 1979). The agreement data was processed through the product analysis division of the Observational Studies Data Analysis System (Clinical Project Research Team, 1978).

Results

The analysis of the "statistical agreement" data, indicated no meaningful difference in the diagnostic problem-solving performance of clinicians on SIMCASES or PAPERCASES. The PAPERCASES proved to be a more efficient and equally as sensitive an instrument as SIMCASES for observational study of reading diagnosis.

Implication

The major implication of this study was that PAPERCASES have the potential to be used as a creative, portable, inexpensive simulation instrument for observing, training, and evaluating reading specialists in the process of diagnosing reading disabilities.

Furthermore, this study made explicit the need for a practical model of the diagnostic process. Recurring behavioral patterns noted among the subjects for this study suggested that the diagnostic strategy employed by these subjects proceeded in a haphazard or random manner. Accordingly, individuals performing as reading diagnosticians need to be provided with a general framework or set of principles from which to proceed in order to increase their consistency and accuracy of diagnosis.

In memory of my father

Michael Colello

who was so proud of his daughter

the school teacher

ACKNOWLEDGMENTS

For their indispensible help, I am indebted to my doctoral committee whose individual and collective efforts have made the completion of this dissertation possible.

My appreciation is extended to Dr. Byron H. Van Roekel who guided my professional development through a course of study which strongly supported my own growth and development in the fields of reading and educational psychology.

I am especially grateful to Dr. George B. Sherman who taught me how to diagnose reading disabilities, how to teach future teachers about the reading process, and who was my mentor and my friend.

I am also grateful to Dr. John F. Vinsonhaler who gave me the opportunity to work with the Clinical Studies Project through which I learned more about the research process than I ever could have learned writing a dissertation on my own.

A special word of thanks to Dr. Glen O. Cooper who reminded me of the real world of public school education when I was emersed in coursework and challenged me to integrate theory with practice.

In addition to my committee members, I wish to extend my appreciation to the Clinical Studies Project staff for supportive assistance, especially Ruth Polin, my computer consultant and Annette Weinshank, my research cohort.

I remain forever grateful to a group of friends whose inspiration and concern I will always treasure: Lynda Sorenson, James Mullin, Lupi and Richard Hamersma, Rose and Sam Hayden, Lynn and Kaela Reynolds, John and Sharon Aycock and Donna Edison.

The friendship of Elaine Stephens, Luthene Chappell, and Joel VanRoekel, my fellow doctural students, as well as many professors, in particular Dr. Gerald Duffy, and secretaries Jacqueline Babcock and Virginia Wiseman, helped make my stay at Michigan State University enjoyable.

Also, I thank my parents, Michael and Phyllis Colello, who saw the wisdom of education for their daughter and sacrificed greatly that I might have it. Most of all, I thank my husband, Jim, for so much love and understanding throughout the seemingly endless process of finishing this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

The International Reading Association (1979) recently specified the minimum standards for the training of four types of specialists in reading: the special teacher of reading, the reading clinician, the reading consultant, and the reading supervisor. In all four role classifications, skill in diagnosing reading disabilities was stipulated as essential. Currently, according to Guthrie (1976), 252 colleges and universities in the United States offer graduate degrees in reading instruction. Without exception, these institutions include, as a requirement for the degree, coursework in the diagnosis of reading difficulties and usually, in addition, a clinical field experience.

In public education, classroom teachers refer millions of children every school year to the many thousands of reading specialists who staff federally funded reading programs such as Title I. Referrals are made on the assumption that the reading specialist will, through the application of a variety of measures,

identify the components of the reading process which are causing the child's reading problem. On the basis of the reading specialist's diagnosis, decisions are made as to who is to be placed in the supplementary remedial reading class, who is to be tutored by a teacher's-aide, who is to practice reading with the parent volunteer, who is to receive more intensive instruction in a specified reading skill, and, in general, who is to receive extra attention and time on the task of reading. Because diagnosis provides the starting point for remediation, teacher-educators as well as practitioners in the field of reading consider expertise in diagnosis to be a basic skill requirement of reading specialists.

While there is a consensus in theory and in practice that diagnosis is a core concept in the field of reading, what constitutes a diagnosis is open to divergent points of view and what the optimum conditions and procedures for collecting and interpreting data are is largely unknown. Until quite recently, research studies on the diagnostic process have been nonexistent (Spache, 1968). As Shulman and Elstein (1973) have observed: "Research typically slights the problem of how teachers think about their pupils and instructional problems; it concentrated instead on how teachers act or perform in the classroom" (p. 3).

Since its founding in April of 1976, the Institute for Research on Teaching (IRT), a research center funded by the National Institute of Education and located at Michigan State University, has been studying the decision-making patterns of experienced teachers. In particular, the Clinical Studies Project within the IRT has addressed the question: "How do experienced reading clinicians approach the diagnosis of children with reading problems?" The study described here--an extension of the research previously initiated by the Clinical Studies Project--was designed to seek another way to obtain information of value in answering the previous question as well as to pursue another question, namely: "What is the more efficient way to initially collect information about the diagnostic problem solving behavior of reading clinicians and to eventually train reading specialists in the process of diagnosing reading disabilities?"

Purpose of the Study

Research to date, within the Clinical Studies Project, on how experienced reading clinicians approach the diagnosis of children with reading problems has been conducted exclusively through the utilization of the SIMCASE--an observational instrument for data collection. A SIMCASE is a device which provides an operating model

or replicia of real world processes. It represents a child with a commonly occurring reading problem and attempts to replicate the clinician-client or reading specialist-student interaction which occurs during the process of diagnosing a reading problem. These simulations allow for observation of diagnostic problem-solving behavior outside of the field setting in which this process is typically performed. The SIMCASE is contained in a file box and consists of six different categories of information about a case of reading disability: (1) test scores, (2) test booklets, (3) test directions, (4) test description, (5) audio-recordings, and (6) examiner's comments. Information for clinical problem solving is retrieved from the SIMCASE by the administrator or observer upon the request of the subject or reading clinician.

That the SIMCASE should prove to be a reasonably effective device for observing clinical problem solving behavior within the Clinical Studies Project was not unexpected because other research has repeatedly demonstrated the appropriateness of simulations for observational studies (Elstein, Shulman, and Sprafka, 1978). However, the tedium inherent in the individual administration of the instrument is potentially a significant limitation in the conduct of research and the training

of reading clinicians, especially when larger numbers of subjects or trainees are involved.

The purpose of this study was to develop and to test the usefulness of an adapted format of the SIMCASE, the intent of which was to mitigate the temporal demands of the individually administered observational instrument and yet retain the sensitivity of the SIMCASE in the observation of diagnostic problem solving performances. Three factors which influence the efficacy of simulations as observational instruments were carefully examined: (1) cost in terms of materials and remuneration to subjects, (2) size in terms of pages, and (3) administrative time per subject or trainee. Careful consideration of these effecting factors revealed that cost and convenience were not amenable to major adaption without significant losses in range, scope, and representation. The remaining option, then, was an alteration which would reduce the administration time requirement per subject or trainee. Accordingly, the SIMCASE was adapted to allow for group-administration, and the resulting instrument was named PAPERCASE.

An observer can administer a PAPERCASE to a large number of subjects within a two hour period as opposed to administering a single SIMCASE to one subject in the same time period. Furthermore, the PAPERCASE is independent

of the observer; it is limited to a subject retrieving information from the pages of a booklet.

The PAPERCASE included four categories of information about a case of reading disability: (1) test scores, (2) test booklets, (3) test directions, and (4) test description with the pages of examiner's comments and the audio-recordings excluded from the PAPERCASE as a minor adaptation in the size of this observational instrument. All audio-recordings and examiner's comments were judged to be nonessential to diagnosing the reading problem and were omitted from the PAPERCASE. This omission represents from 14 to 35 pieces of information or cues depending on the particular case of reading difficulty. Thus, the PAPERCASE contains roughly 20 percent fewer cues than the SIMCASE format.

Rationale for the Study

In comparison with the SIMCASE simulation, the PAPERCASE may prove to be a more efficient but equally sensitive instrument for use in observing diagnostic problem-solving performance. Furthermore, in addition to using PAPERCASES as an observational instrument directed toward understanding the process of diagnosis, the PAPERCASES have high potential as training exercises intended to teach the process of diagnosis. PAPERCASES

may offer a creative, very portable, and inexpensive way of vitalizing graduate and in-service teacher education by fusing theory and practice into efficient learning of diagnostic skills.

Hence, individuals training to become reading specialists may learn about a number of reading problems without having to resort to clinical experience to reinforce what they have learned. Moreover, even when clinical experience is available and the opportunity is there to actually diagnose a real child's reading problem, staffing problems often pose severe limitations on the ability of the clinical faculty to observe a student's performance and provide the student with the feedback that is necessary to maximize the development of skill in diagnosis (Van Roekel, personal communication). The PAPERCASE is an alternative to the direct observation of an individual diagnosing a reading problem. With the addition of a feedback component, the PAPERCASES could provide students with the opportunity to practice solving a range of reading problems which closely approximate, in breadth and complexity, the problems which they will encounter as reading specialists in the real world.

The Research Objective

In designing the study it was recognized that there are a number of researchable questions of potential

interest with respect to the proposed new instrument. The decision as to which questions would be asked, and which would be deferred for subsequent investigation, was largely decided by this author's conception of the research strategy which would make the most immediate, direct contribution to the ongoing Clinical Studies Project. This strategy suggested the following research objective: Evaluate the usefulness of PAPERCASES as an observational instrument by comparing the diagnostic performance of clinicians on PAPERCASES with performance on SIMCASES. This research objective questions: can the PAPERCASE be substituted for the less efficient SIMCASE as an instrument which is equally effective in collecting data on diagnostic problem-solving performance? The answer to this question depended upon the consistency of the clinicians' performance on the two instruments.

Theoretical Basis of the Study

When one considers the process of reading diagnosis, one is faced with describing the very complex and cognitive behavior of problem solving. No intelligible description of phenomena as abstract as problem solving is possible without a sound theoretically based framework from which to study the nature of this behavior.

The Inquiry Theory of Clinical Problem Solving formed the theoretical basis for this study. Similarities

drawn by Elstein, Shulman, Vinsonhaler, and others (1977) between Clinical Problem Solving which was initially developed in a medical mode and the behavior of reading clinicians which is in an educational mode, lends strong credibility to the application of the Inquiry Theory to Reading. Feature by feature, the authors illustrate the correspondence in process between the medical-clinical task of diagnosis and the reading-clinical task of diagnosis. In summary terms, both clinicians are concened with alleviating problems presented to them by individuals who may be identified as the patient, the client, or the student in difficulty. More importantly, both clinicians as practitioners are considered to informally proceed through a cycle of data collection--hypothesis generation--cue interpretation--hypothesis verification in the process of problem solving.

Assumptions of the Inquiry Theory

Exploring the nature of Clinical Problem Solving in Reading through observational studies, Vinsonhaler, Wagner, and Elstein (1977) have made explicit the behavioral domain addressed by the Inquiry Theory. The behavioral domain of this theory is known as the Clinical Encounter which encompasses the events occurring when a Clinician (e.g., physician, reading specialist) interacts with a Case (e.g., a patient, a student) in order to reach

a diagnostic and/or treatment decision about the presenting problem (e.g., child has a temperature, child cannot read his textbooks). Thus, the Clinical Encounter includes: (1) the Clinical Case, (2) the Clinician, and (3) the Clinical Interaction. These three components in the Clinical Encounter are illustrated in Figure 1 and the direction of the interaction is indicated by arrows.

The Clinical Encounter is the basis of the first assumption of the Inquiry Theory which states: "The behavioral domain addressed by the theory involves a clinician, a case or patient, and an interaction which yields a decision on the Diagnosis (the state of the case) and the Therapy (how this state can be improved)" (Vinsonhaler, 1979). The second assumption relates to the clinical case, and states: "Important problem solving behaviors of children can be elicited by simulated cases" (Vinsonhaler, 1979). The use of simulated cases is based on the educational principal that problem solving skills can best be "learned by doing" (Dewey, 1963; Bruner, 1966; Gagne, 1971).

The third assumption, presented in Figure 2, focuses on the Clinician and describes how Clinical Problem Solving occurs: Clinical Problem Solving is determined probabilistically by the interaction of (1) Clinical Memory, (2) Clinical Strategy, and (3) the Case (Vinsonhaler, 1979).

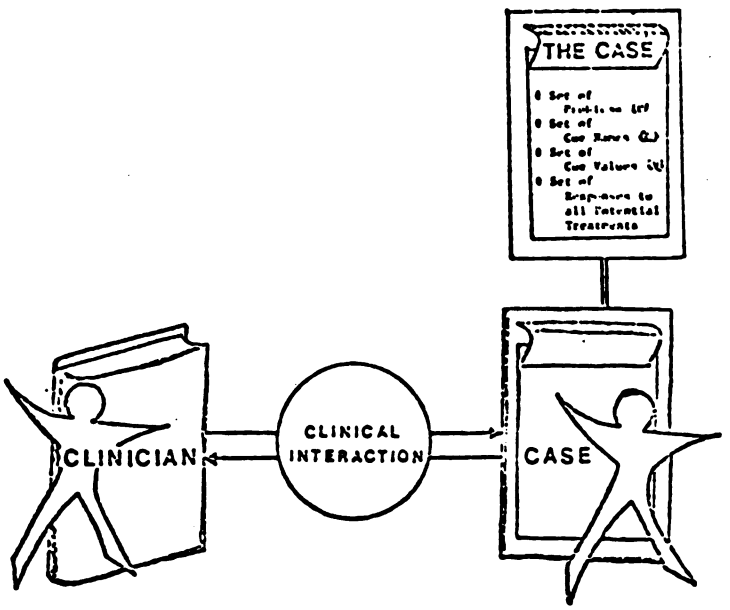

Figure 1.--The Clinical Encounter.

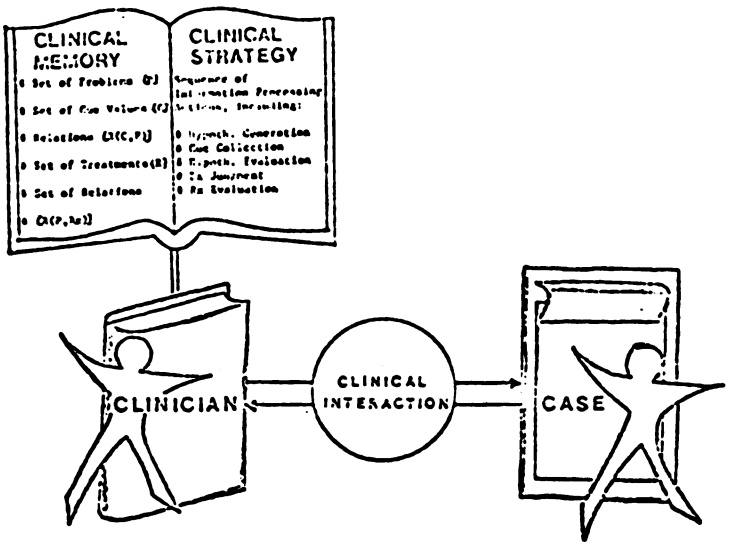

Figure 2.--Clinical Problem Solving.

Clinical Memory consists of sets: (1) problems, (2) cues, and (3) treatments and the relationships among them. An example of Clinical Memory in the context of reading diagnosis is:

1. Problem - Child does not attempt to answer inference questions.
2. Cue - Iowa Tests of Basic Skills (ITBS) Reading Subtest, all inference questions unanswered.
3. Relating problem and cue - Item analysis of errors on the ITBS - Reading Subtest suggest the likelihood that a comprehension problem is high.

Clinical Strategy consists of the mental tasks performed by the Clinician which translate memory into action. These tasks mainly involve information-gathering and information-processing as the clinician makes decisions about diagnosis and remediation (Vinsonhaler, Wagner, and Elstein, 1977).

The Research on the Inquiry Theory of Clinical Problem Solving initially conducted by the Medical Inquiry Project form 1969 through 1973 at Michigan State University in the context of medical education has been continued over the past three years by the Clinical Studies Project in the context of teacher education. The Clinical Studies Project is applying the original concepts developed in medical diagnosis and treatment to the field of reading diagnosis and remediation.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

"Concepts are properties of organismic experience--more particularly, they are the abstracted and often cognitively structured classes of 'mental' experience learned by organisms in the course of their life histories. . . . Within a given community there will be a high degree of commonality in the concepts recognized and attained, in the sense that there will be relatively high agreement among people as to the attributes that are criterial for a given concept" (J. B. Carroll, 1964, pp. 180 and 185). The purpose in this chapter is to mark off the boundaries of the concept of diagnosis in the field of reading-education by attending to selected attributes of diagnosis described in the literature.

The Nature of Diagnosis

What generic observations have been made about diagnosis in the literature? Description of the nature of diagnosis is very limited. To begin with, the Random House Dictionary of the English Language generally

defines diagnosis as an answer or solution to a problematic situation. According to Della-Piana (1968), "to diagnose is to determine the nature of a process by examining in some detail the differences between the functioning of various parts of the process" (p. 3). Dechant (1968) stated that the heart of diagnosis is an intelligent interpretation of facts, it is an inference from performance.

The Institute for Research on Teaching (IRT) has been studying teaching as diagnosis. The IRT has broadly defined diagnosis as a decision-making or problem-solving process in which a clinician interprets information about individual students or a class as a whole. The clinician combines information about students and classes, as well as information from the educational research literature, with his/her own expectations, attitudes, beliefs, and purposes. Based upon all this processing of information, the clinician then responds, renders decisions, and regroups to begin again (Cruickshank and Kennedy, 1977).

The description of diagnosis most closely related to the IRT's point of view comes from H. L. J. Carter (1970) who specified the following four acts as integral to the process of diagnosis:

1. Identify the problem and possible causal factors.
2. Assume and reject hunch after hunch until one can be accepted tentatively.
3. Discover possible determinants and explain consequential relationships.
4. Predict that with treatment the disability will be overcome (p. 20).

Similarly, Spache (1976) collaborating with R. W. Prouty described diagnosis as "a continuous process of proposing hypotheses, testing them by teaching strategies and referring or discarding them. . . . As we test and as we begin instruction that seems relevant in terms of our first impressions, we must constantly observe the pupil's behavior response to the approach we are using and its apparent impact upon his development" (p. 9).

While theoretically, diagnosis has been generically described as a problem-solving process it is practically described in the literature in terms of its purposes.

The Purpose of Diagnosis in Reading

In 1935 Bruecker discussed diagnosis as techniques by which one discovers and evaluates the strengths and weaknesses of an individual. Much later, Smith and Dechant (1962) elaborated on this purpose:

> Diagnostic procedure begins with a study of the child's instructional needs based on the expectancies of his chronological age, mental age, and grade placement. We seek to discover why he reads as he does, what he can read, and what he does read successfully. We need to know if he is having problems in reading and, if so, what they are and what are their causes. We wish to know his general abilities and his reading potentiality and we must identify causal factors that have retarded his reading development. In short, we must know his strengths and weaknesses (p. 408).

Durkin (1970) has written about the positive and negative overtones of reading diagnosis in contrast with medical diagnosis. She noted that medical diagnosis has a negative overtone since medical personnel are usually trying to learn what is wrong with a patient but reading diagnosis has both a positive and a negative dimension. It is just as concerned about what a child knows and can do as it is about what he does not know and cannot do (p. 402).

Focusing on the negative dimension, H. L. J. Carter (1970) stated diagnosis is an explanation of an individual's maladjustment in reading. In the study of a disabled reader, the teacher and clinician are concerned with cause which precedes an event called the effect. The purpose is to determine why the individual is disabled, what went wrong (p. 17).

Specifying the purpose of diagnosis as the identification of weaknesses and strengths is consistent with what Strang (1969) identified as the first level

of diagnosis. On the first or surface level, an effort is made to describe reading performance--strengths and weaknesses in vocabulary, work recognition, sentence and paragraph comprehension, and related abilities. Going beyond the surface, other authorities in the field have set forth additional purposes for reading diagnosis.

Levels of Diagnosis in Reading

Diagnosis of reading disabilities may be made on different levels of comprehensiveness, psychological depth, and competence (Strang, 1964, p. 4). Monroe (1937) explained that the diagnosis of reading disabilities should contain two types of analysis: (1) descriptive and (2) causative. In the descriptive analysis the examiner details the nature of the child's reading disability based upon subjective observation and objective test data. In the causative analysis the examiner investigates five areas: (1) constitutional, (2) intellectual, (3) emotional, (4) educational, and (5) environmental, which may be contributing to the problem (p. 14).

Extending Monroe's (1937) two levels of analysis to four, H. L. J. Carter (1970) stated diagnosis proceeds from (1) identification of difficulty, to (2) classification of disability, to (3) determination of specific needs, and at the highest level (4) detection of causal factors underlying the individual's disability (p. 18).

Bond and Tinker (1973) discussed diagnosis in terms of successive screenings in which the more complex and subtle cases of reading disability are retained for further analysis at more in-depth levels. Referring to Brueckner and Bond (1955), Bond and Tinker (1973) believed some cases of reading disability must be carried through three levels: (1) general diagnosis, (2) analytical diagnosis, and (3) case-study diagnosis. General diagnosis is made by studying test results in order to locate general areas of weakness. Analytical diagnosis identifies specific strengths and weaknesses and indicates skills and abilities wherein the child's weakness lies. The case-study diagnosis analyzes information on mental capacity, vision, hearing, physical characteristics, adjustment to reading, and environmental factors bearing upon the child. Again, the level of diagnosis reached depends upon the characteristics of the case (pp. 168-171).

Rutherford (1972), addressing the classroom teacher, asserted that teachers accept a range of four explanations of a reading disability which proceed from general to specific. At level one difficulties are classified in broad categories of causes: level two applies special terms to the problem; level three describes certain types of general, overt student behavior; and level

four is prescriptive in terms of what the child needs to learn (pp. 51-53).

Strang (1968) further analyzed the breadth of possible difficulty in reading and has ordered seven levels of reading diagnosis. On the first level information is obtained on student performance. On the second level the student's reading behavior is observed. On the third level an attempt is made to analyze the student's reading process as opposed to the student's performance. On the fourth level mental capability is investigated. The fifth level involves clinical analysis of personality traits and values. The sixth level involves an examination of possible pathological conditions, such as brain damage. The approach at the seventh level is to ask the reader to describe his reading process, termed introspective reports (pp. 4-6).

In sum, the numerous levels of diagnosis suggest what may constitute this problem-solving behavior of diagnosis and the possible level of specificity that may be necessary in a diagnosis depending on the complexity of the case. Of course, the reading specialist's problem is to determine just what is, and is not, necessary. But, necessary for what? The widely acknowledged answer is--for remediation. According to such authorities as Bond and Tinker (1976), Ekwall (1976), and Spache (1976),

reading specialists should gather only enough initial diagnostic information to begin a program of remediation.

Diagnosis as a Prerequisite to Remediation

Several well-known authors of textbooks on the subject of diagnosis and remediation repeatedly position diagnosis as a prerequisite to remediation. To quote Monroe (1973), "To be effective, remedial instruction in reading must be preceded by careful diagnosis" (p. 359). Writing as far back as 1922, Gray stated: "After diagnosis has shown the kind of instruction that is needed, the remedial program should be carefully planned" (p. 374). In the words of Dechant (1968), "Diagnosis is complete only when remediation occurs" (p. 6). As Della-Piana (1968) puts it, "The major purpose of diagnosis is to gather information that may be helpful in making treatment decisions" (p. 3). In Spache's (1976) terms, "Diagnosis is pragmatic and directly related to remedial practice" (p. 9).

Other authors publishing in the field have reiterated the same position. E. C. Kennedy (1971) stated: "The real purpose of educational diagnosis is to secure specific information about a pupil which will enable teachers to plan for direct and appropriate instruction" (p. 97). Also, R. M. Wilson (1977)

believed that "Regardless of the educator's professional position, diagnosis is essential in formulating a remedial program which will be both effective and efficient" (p. 15). A final reiteration from Karlsen (1976), "Diagnosis is not simply analysis of basic causes of reading problems. It is oriented toward the future and is most effective when it helps the teacher arrange meaningful and efficient learning experiences that will help each student become a skillful reader" (pp. 147-148).

Measures of Teacher Knowledge of Reading

The agreement among authorities in the field of reading-education that diagnosis precedes remediation is based on the assumption that reading clinicians and classroom teachers know how to diagnose and remediate reading difficulties. However, little research evidence exists to confirm this assumption (Gil, Vinsonhaler, and Sherman, 1979). Research studies have been limited to measuring clinician knowledge of reading. A brief review of these studies focusing on the instruments developed to measure teacher knowledge of reading and the performance of teachers on the instruments follows.

Schubert (1959) was interested in finding out elementary and secondary teacher's knowledge of structural

and phonetic analysis. He administered an informal ten question quiz to 80 elementary teachers and 42 secondary teachers and reported that a substantial number of them did not possess knowledge of certain principles of word analysis.

A sixty-item multiple-choice test on phonetic-generalizations was developed by Aaron (1960) and administered to 104 persons with one or more years of teaching experience and 189 persons with no teaching experience. Results indicated that very few subjects were well-grounded in phonics principles.

Spache and Baggett (1965) used a modified version of Aaron's test with graduate students and inservice teachers pursuing graduate work and found that they were generally weak in the areas of phonics and syllabication.

Ramsey (1962) and Browman (1962) both developed tests to determine the extent to which teachers possessed knowledge of basic skills in reading. Durkin's (1964) test called the Phonics Test for Teachers was designed for use in reading methods courses to help students identify what they know and what they do not know about phonics.

An instrument which covered rather broad areas of reading was developed by Wade (1960) to measure such skills as "diagnosing and correcting phonic and syllabication errors." His instrument included an audio-tape

and paper-and-pencil questions. He administered his test to students, teachers, and reading specialists and found, as expected, the students achieved the lowest and the reading specialists achieved the highest.

The Inventory of Teacher Knowledge of Reading was recently developed by Artley and Hardin (1975). This test contains 95 multiple-choice items and covers the following areas: the reading act, preparation for reading, word identification, comprehension and critical reading, reading in the content areas, reading interests and tastes, and corrective procedures. However, factor analysis indicated that the seven areas from which the items were drawn were not identifiable as discrete areas. Kingston and his associates (1975) also analyzed the results of the Inventory of Teacher Knowledge of Reading. Kingston's factor analysis also failed to reveal seven components of this Inventory.

The measures of teacher knowledge of reading by Schubert (1959), Aaron (1960), Spache and Baggett (1965), Ramsey (1962), Browman (1962), Durkin (1964), Wade (1960), and Artley and Hardin (1975) are all possible instruments for estimating some of the categories of information a clinician brings to the process of diagnostic-problem solving. In this respect, these measures relate to "clinical memory," a component of the Inquiry Theory of Clinical Problem Solving which forms the theoretical

basis of this study. As initially presented in Figure 2 of Chapter I, the third assumption of this theory is that clinical problem-solving is determined probabilistically by the interaction of (1) clinical memory, (2) clinical strategy, and (3) the case (Vinsonhaler, 1979). Clinical memory includes the background of experience and information a clinician brings to a case with which to make decisions about the case and these measures of teacher knowledge may partially describe a clinician's clinical memory.

Measures of the Diagnostic Ability of Teachers

Burnett (1961) developed a test to measure the diagnostic ability of teachers. His test consisted of problems to measure five levels of operation in diagnosis. The first level problems required the examiner to pick critical information from a pool of data; the second level problems required selecting a means of securing additional data; the third level required the interpretation of data; the fourth level required recommendations for improving instruction; and at the fifth level, the examiner was required to re-evaluate his fourth level recommendations. Burnett found that reading specialists significantly outscored experienced teachers and experienced teachers outscored the undergraduate student.

The mean differences were significant beyond the .01 and .05 levels.

Emans (1965) assessed the diagnostic ability of teachers enrolled in a clinical practicum in reading remediation. For an hour a day over five weeks, 20 teachers provided individualized reading instruction to two children experiencing some difficulty in reading. At the end of the five week period the teachers were asked to rank 15 reading skills in the order in which their children needed help on them. The teacher's rankings were compared with the rankings of the skills as indicated on the Reading Diagnostic Test by Arthur Gates. The correlations between the teachers' judgments and the scores on the test were very low. Emans suggested that individualized reading programs were doomed from the start if teachers were unable to determine the children's reading needs.

Observations of the Diagnostic Skill of Reading Clinicians

The Clinical Studies Project has developed simulation instruments rather than tests to observe and objectively describe performance of clinicians resulting from the interaction of clinical memory, clinical strategy, and a case. The research approach of this project has been to study the "wisdom of the practitioner" with individual teachers recognized as having expertise in

reading diagnosis not only participating as clinical research subjects but also as sources of insight regarding their own behavior and functioning when presented with simulated cases of reading difficulty. To date, the Clinical Studies Project has conducted a series of six observational studies as part of a systematic program of research. All of the observational studies have shared a common theoretic and methodological base which included the utilization of SIMCASES as the instrument in data collection.

In very global terms, the findings of the first observational study, OS77.1, were that performance on the SIMCASES was inconsistent. That is, a problem diagnosed one way by one reading clinician was diagnosed differently by another. In addition to not agreeing with each other when diagnosing and remediating the same case, the reading clinicians disagreed with themselves when diagnosing different versions of the same problem (original SIMCASE and replicate SIMCASE). However, the combined diagnosis of several clinicians or group diagnosis appeared to be more consistent than the diagnosis by an individual reading specialist.

Considering the additional observational studies, all indications are that a second observational study, OS78.3, being conducted by L. Hoffmeyer and termed a "conceptual replication," will verify the findings of

the first observational study, OS77.1. A third study, OS78.2, focusing on the diagnostic performance of classroom teachers as opposed to reading specialists, was conducted by D. Gill. His investigation centered on the identification of the classroom teacher as a diagnostician who is in the position to observe children's reading behavior on a continuous basis and diagnose and treat reading problems at their inception, long before the problem becomes severe and is brought to the attention of the reading specialist. Gil's outstanding general results were very similar to the findings of the first observational study, OS77.1. Classroom teachers were inconsistent with each other when diagnosing and remediating the same case. "The most frequently mentioned diagnostic judgments differed from case to case. . . . On the average, teachers in this study showed very limited agreement with each other on diagnostic judgments and cues collected for a given case" (Gil, 1979, pp. 104 and 105).

A fourth study, OS78.1, now in progress by J. Van Roekel is investigating the problem solving behavior of both teachers trained in the field of Reading and teachers trained in the field of Learning Disabilities. According to Van Roekel, his study was designed with the following purposes in mind:

> First, to identify the types of measures typically utilized by clinicians representing the fields of reading and learning disabilities when they interact with identical cases of education difficulty. Second, to determine if there are systematic differences between these two groups of clinicians with regard to the type of data collected, the amount of data collected, the degree ot which that information is used in a diagnostic summary, and the diagnostic conclusions drawn by clinicians representing the two fields. Third, to compare problems solving processes and diagnostic products of reading clinicians and learning disabilities resource teachers employed in public schools with a group of senior clinicians. Finally, to explore the application of the Inquiry Theory of Clinical Problem Solving to the diagnosis of reading and learning problems (Gil, Hoffmeyer, Van Roekel, Vinsonhaler, and Weinshank, 1979, p. 32).

A fifth study, OS79.2, raised the question: "What is the relationship between diagnosis and remediation?" A. Weinshank is systematically investigating a corollary to the Inquiry Theory of Clinical Problem Solving which states: Problems and treatments are associated in probabilistic fashion such that given a problem some treatments would be more likely to be selected than others." This corollary, termed problem-treatment association (PTA), was previously tested by Weinshank (1978) in a pilot study. The pilot study analyzed the therapeutic decisions contained in the remediation plans written by the reading diagnosticians who were the subjects in the first observational study, OS77.1. Her major findings were than diagnostic statements outnumbered treatment statements by a margin of from 2:1 to 7:1; the relationship between problems stated in the diagnosis and

treatments proposed in the remediation was moderate; clinicians who agreed with themselves more on the diagnosis of a case also agreed with themselves more on the remediation for that case; and clinicians agreed on the use of a core subset of treatments across almost all of the cases.

This present study is a sixth observational study, OS79.1, the methodology and results of which are discussed in the following chapters.

CHAPTER III

DESIGN OF THE STUDY

Introduction

The construction of a PAPERCASE included a developmental phase and an experimental phase. The development of the PAPERCASE will be outlined first followed by a description of the agreement statistics used in the experiment to compare the diagnostic performance of clinicians on SIMCASES with performance on PAPERCASES.

Development of the SIMCASES

The SIMCASES are a product of previous research in the Clinical Studies Project of the IRT. Each SIMCASE is presented in a metal file box and represents actual cases of male students in second through tenth grade who have a reading disability. Each SIMCASE includes six categories of information on the case: (1) test scores, (2) test booklets, (3) test direction, (4) test description, (5) audio-recordings, and (6) examiner's comments. Three original SIMCASES representing actual cases of reading disability were used as observational instruments

and three replicate forms of these self same original SIMCASES were used in this study.

Development of PAPERCASES

Three PAPERCASES were constructed by this researcher from the replicate forms of the three original SIMCASES. Each PAPERCASE is contained in a booklet and includes four categories of information on each case: (1) test scores, (2) test booklets, (3) test directions, and (4) test descriptions.

Subjects

The subjects of this study were 12 graduate students who studied Diagnosis of Reading Difficulty, Education 830E, at Michigan State University under Dr. George Sherman, within three past academic terms (Winter, Spring, and Fall, 1978) and earned a final grade of 4.0 thereby demonstrating to Dr. Sherman initial mastery of the requisite skills for diagnosing reading problems. A further requirement of the subjects for participation in the study was that they be certified classroom teachers. This sample was chosen because:

1. This group of graduate students had similar amounts of knowledge about diagnosis and similar practical experience in diagnosing reading problems and writing

diagnostic case studies through the course requirements of Education 830E.

2. This group of graduate students was generally familiar with the diagnostic testing materials included in the observational instruments for this study. Many of the same testing materials were used in Education 830E to collect information on a child with a reading problem.

3. This group of graduate students was judged to be more accessible than a group of expert reading diagnosticians. Previous observational studies nearly exhausted the local population of experts in the field.

Sample Selection

The graduate students were contacted by telephone and asked to participate at a rate of $9.00 per hour as subjects in a study of diagnostic problem solving. Subjects were contacted in random order and randomly assigned to the three cases of reading disability and the two observational instruments.

Presentation of the SIMCASES and PAPERCASES

Table 1 illustrates the manner in which SIMCASES and PAPERCASES were presented to the subjects. Each of the twelve subjects were randomly assigned to a case and a replicate form of the same case. Three original

Table 1

Presentation of Cases and Instruments

Subject	Case	First Session Instrument - Form	Second Session Instrument - Form
501	I	SIMCASE - original	SIMCASE - replicate
502	I	SIMCASE - replicate	SIMCASE - original
503	II	SIMCASE - original	SIMCASE - replicate
504	II	SIMCASE - replicate	SIMCASE - original
505	III	SIMCASE - original	SIMCASE - replicate
506	III	SIMCASE - replicate	SIMCASE - original
507	I	SIMCASE - original	PAPERCASE - replicate
510	I	PAPERCASE - replicate	SIMCASE - original
508	II	SIMCASE - original	PAPERCASE - replicate
511	II	PAPERCASE - replicate	SIMCASE - original
509	III	SIMCASE - original	PAPERCASE - replicate
512	III	PAPERCASE - replicate	SIMCASE - original

SIMCASES, three replicate forms of the original SIMCASES, and three PAPERCASES constructed from the replicate forms of the original SIMCASES were presented to the subjects.

Four subjects were assigned to the same case of reading disability. Subjects 501, 502, 507, and 510 diagnosed Case I in a total of eight sessions. Subjects 503, 504, 508, and 511 diagnosed Case II in a total of eight sessions. Subjects 505, 506, 509, and 512 all diagnosed Case III in a total of eight sessions. In sum, there were three different cases of reading disability with four subjects assigned to a case and each subject participating in two sessions spaced a week apart generating a total of eight sessions per case of reading disability.

The design technique of counterbalancing was applied to the presentation to control for the extraneous variables of amount of practice and fatigue. The general principle of this technique may be stated as: Each condition must be presented to each subject an equal number of times and each condition must occur an equal number of times (McGuigan, 1968). Since subjects were assigned to two observational sessions, the possibility existed that they might learn the task during the first observational session and apply what they learned to the second observational session. Within the second observational session a certain amount of improvement in the subjects' performance could be due to practice, conversely, a

certain decrement in performance could be due to fatigue. According to McGuigan (1968): "The method of counterbalancing attempts to distribute these effects equally to all conditions. Hence, whatever the practice and fatigue effects, they presumably influence each condition equally since each condition occurs equally often at each stage of practice" (p. 134).

Procedure for SIMCASE Data Collection

Data was collected from subjects assigned to SIMCASES in individual sessions lasting about two hours. Each SIMCASE session began with the subject completing information forms, the examiner administering a set of standardized instructions, and then continued with the following systematic set of procedures:

1. The examiner provided an overview of the session (see Appendix A for Instructions).
2. The subject was given a Cue List which specified the six types of Cues (information) available on request (see Appendix B for the Cue List for each Case).
3. The examiner explained how to use the Cue List.
4. The subject practiced using the Cue List to request information about a sample SIMCASE.

5. The examiner introduced the SIMCASE assigned to the subject by presenting a sketch of the child, a statement of the presenting problem, and a taped interview with the child.
6. The subject was given a maximum of 30 minutes to use the Cue List to request information about the SIMCASE and reach a diagnostic decision.
7. The subject was given a maximum of 20 minutes to write a diagnosis for the SIMCASE.
8. The subject was given a short break.
9. The subject was given a maximum of 20 minutes to write a remediation plan for the SIMCASE.
10. The examiner explained to the subject procedures for transferring the hand-written diagnostic statements to a standardized Diagnostic Checklist (see Appendix C for the standardized Diagnostic Checklist).
11. The subject practiced transferring his/her own handwritten diagnostic statements to the standardized Diagnostic Checklist.
12. The subject transferred all written statements to the standardized Diagnostic Checklist.

13. The examiner explained to the subject the procedure for transferring written remediation statements to a standardized Remediation Checklist.
14. The subjects practiced transferring his/her own written statements to the standardized Remediation Checklist.
15. The subject transferred all written statements to the standardized Remediation Checklist.
16. End of session.

Procedures for PAPERCASE Data Collection

Data was collected from the subjects assigned to the PAPERCASES in small group sessions lasting about two hours. Each PAPERCASE session began with subject completing information forms and the examiner administering a set of standardized instructions to the group followed by the same set of systematic procedures utilized for SIMCASE data collection with the following initial five modifications for this booklet format:

1. PAPERCASE booklets were passed out to the small group of subjects.
2. The examiner provided an overview of the session.

3. The subjects were instructed to open their booklets to the page entitled Instructions (see Appendix D for Instructions).
4. The examiner read three typewritten pages out loud to the subjects and the subjects were asked to follow the typed text.
5. The subjects were given a maximum of 30 minutes to use the Cue List in the booklet to collect information on the case and reach a diagnostic decision (see Appendix E for the Cue Lists for each Case).
6. The subject was given a maximum of 20 minutes to write a diagnosis for the PAPERCASE.
7. The subject was given a short break.
8. The subject was given a maximum of 20 minutes to write a remediation plan for the PAPERCASE.
9. The examiner explained to the subject procedures for transferring the handwritten diagnostic statements to a standardized Diagnostic Checklist (see Appendix C for the standardized Diagnostic Checklist).
10. The subject practiced transferring his/her own handwritten diagnostic statements to the standardized Diagnostic Checklist.

11. The subject transferred all written statements to the standardized Diagnostic Checklist.
12. The examiner explained to the subject the procedure for transferring written remediation statements to a standardized Remediation Checklist.
13. The subject practiced transferring his/her own written statements to the standardized Remediation Checklist.
14. The subject transferred all written statements to the standardized Remediation Checklist.
15. End of session.

Observational Studies Data Analysis

The data was analyzed through the computerized statistical system developed by the CLIPIR Research Team (1978) termed Observational Studies Data Analysis System and referred to as OSDAS. Since the study was concerned with the end results of observational sessions or diagnostic performance on SIMCASES and PAPERCASES, the product analysis division of the OSDAS program system was utilized in the data analysis. In addition to the OSDAS Product programs, a general purpose statistical system developed by the Statistical Research Laboratory at the

University of Michigan termed MIDAS was used for all descriptive statistics such as means and standard deviations.

The Independent Variables

The independent variables in this study consisted of the observational instruments, SIMCASE and PAPERCASE, which were different in format and size. The experimenter manipulated the independent variables by randomly assigning different subjects to the two observational instruments.

The Dependent Variables

How the subjects performed on the SIMCASES or PAPERCASES constituted the dependent variables in this study. With twelve clinicians participating in two observational sessions each, there were a total of twenty-four observational sessions. Four measures of diagnostic problem-solving performance were obtained from the sessions conducted on each case of reading disability, Case I, Case II, and Case III.

Proportional Agreement

The first measure determined whether different forms of a case, SIMCASE or PAPERCASE, influenced the Proportional Agreement on (1) diagnostic cues of information collected on a case and (2) diagnostic statements

recorded about a case. Proportional Agreement is a measure of group agreement on cues and diagnosis. This statistic gives an overall indication of the cue and diagnostic categories most frequently employed by a group of clinicians (Vinsonhaler, 1979).

Given a domain of cues or diagnostic statements for a given case, Proportional Agreement is the proportion of clinical sessions in which each cue or diagnostic statement was mentioned. This statistic is bounded by zero and one and was calculated by dividing the number of clinical sessions in which a statement was mentioned by the total number of sessions on a case. For example, if the category "Basic Sight Words - Weakness" was mentioned in four of the eight sessions conducted on a case the Proportional Agreement would be:

$$P.A. = \frac{\text{Number of Sessions in which Clinicians Mentioned the Category}}{\text{Total Number of Clinical Sessions on the Case}}$$

$$P.A. = 4/8 = .50$$

The research question derived from this measure was: Will the group of clinicians assigned to SIMCASES and/or PAPERCASES proportionately agree any more or less with each other on their categorization of (1) cues or (2) diagnostic statements than other groups of clinicians assigned to only SIMCASES in previous studies?

Commonality Score

The second measure determined whether different forms of a case, SIMCASE or PAPERCASE, influenced the Commonality of (1) diagnostic cues of information collected on a case and (2) diagnostic statements reported about a case. The Commonality Score measured the degree to which an individual agreed with the group assigned to the same case.

Given a domain of cues or diagnostic statements for a given case, the Commonality Score compared an individual's cues and diagnosis with that of a group (Vinsonhaler, 1979). This statistic is bounded by zero and one. A value of "X" for a given clinician roughly implied that he/she has included in his/her session on a case X% of those statements most frequently mentioned by the group for that case. For example, a clinician who has a score of ".34" has included in his/her session roughly 34% of those statements mentioned most frequently by the group.

The research question derived from this measure was: Will the cues or diagnostic statements of an individual assigned to the PAPERCASE agree any more or less with the cues of diagnostic statements of the group assigned to the SIMCASE form of the same case?

Inter-Clinicial Correlation

The third measure determined whether different forms of a case, SIMCASE or PAPERCASE influenced Inter-Clinician Correlation of (1) diagnostic cues or information collected on a case and (2) diagnostic statements reported about a case. The Inter-Clinician Correlation measured the agreement of one clinician with another clinician on the same case. This statistic expressed in terms of a phi coefficient compared cues and diagnostic statements of one clinician with those of another clinician.

Two by two contingency tables were prepared and phi coefficients calculated to describe the presence or absence of cues or diagnostic statements made by one clinician in comparison with the presence or absence of cues or diagnostic statements made by another clinician. The cells in the contingency table included the number of statements: (a) present in both encounters, (b) present in the ith but not the jth encounter, (c) present in the jth but not the ith encounter, and (d) statements not present in either encounter (Vinsonhaler, 1979).

Given a domain of cues or diagnostic statements for a given case, the phi correlation is a measure of Inter-Clinician Agreement. One phi correlation was computed for each pair of clinicians on the same case. This statistic is bounded by -1 (statements are in cells

b and c only) and 1 (statements are in cells a and d only), only if the distributions in the marginals are equal. In all other cases the maximum and minimum values will be less than and greater than -1. The form of the calculation is summarized in Figure 3.

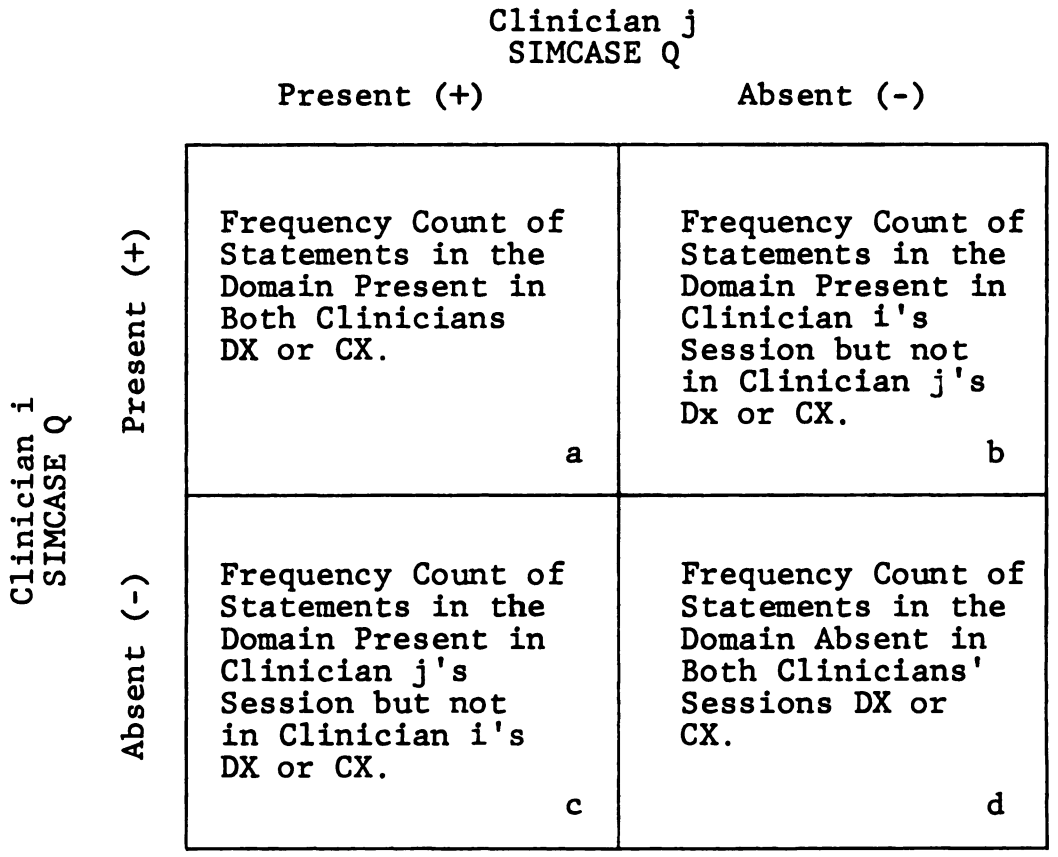

Figure 3.--Form of the Two by Two Contingency Table on Inter-Clinician Correlation.

An example of a completed contingency table on Inter-Clinician Correlation is as follows:

Statements of Clinician i, SIMCASE Q	Statements of Clinician j, SIMCASE Q	Domain Statements
S_1	S_1	S_1
S_2	S_2	S_2
S_3	S_7	S_3
		S_4
		S_5
		S_6
		S_7

		Clinician j: +	Clinician j: -	
Clinician i	+	2	1	3
	-	1	3	3
		3	4	7

The research question derived from this measure of Inter-Clinician Correlation was: Will the Inter-Clinician correlations computed for clinicians assigned to the SIMCASE form of a case be meaningfully different from the Inter-Clinician correlations computed for clinicians assigned to the PAPERCASE form of the same case?

Intra-Clinician Correlation

The fourth measure determined whether different forms of a case, SIMCASE or PAPERCASE influenced Intra-Clinician Correlation of (1) diagnostic cues collected on a case and (2) diagnostic statements reported about a case. Intra-Clinicial Correlation measured a clinician's agreement with him/her self on a case and its replicate. This statistic, also expressed in terms of a phi coefficient, compared the cues or diagnostic statements made by a clinician in one session with the cues or diagnostic statements made by the same clinician in another session with both sessions on the same case.

Two by two contingency tables were prepared and phi coefficients calculated to describe the presence or absence of cues/diagnostic statements by a clinician during the initial session on a case in comparison with the presence or absence of cues/diagnostic statements made by the same clinician when he/she was

presented with a replicate form of the same case during another session. The phi coefficients were calculated following the same procedure specified for the Inter-Clinician Correlation (Vinsonhaler, 1979). The form of this calculation is summarized in Figure 4.

Clinician i SIMCASE Q, Form 2	Clinician i SIMCASE Q, Form 1: Present (+)	Clinician i SIMCASE Q, Form 1: Absent (-)
Present (+)	Frequency Count of Statements in the Domain Present in both Sessions for Form 1 and Form 2 of SIMCASE. a	Frequency Count of Statements in the Domain Present in the Session for Form 2 SIMCASE but not in Form 1 SIMCASE. b
Absent (-)	Frequency Count of Statements in the Domain Present in the Session for but not in Form 2 SIMCASE. c	Frequency Count of Statements in the Domain Absent in Both Sessions for Form 1 and Form 2 of SIMCASE. d

Figure 4.--Form of the Two by Two Contingency Table on Intra-Clinician Correlation.

An example of a completed contingency table on Intra-Clinician Correlation is as follows:

Statements of Clinician i, SIMCASE Q, Form 1	Statements of Clinician i, SIMCASE Q, Form 2	Domain Statements
S_1	S_1	S_1
S_2	S_2	S_2
S_3	S_7	S_3
		S_4
		S_5
		S_6
		S_7

Simcase Q, Form 2 \ Simcase Q, Form 1	+	-	
+	2	1	3
-	1	3	4
	3	4	7

Summary

The four dependent measures were all tested by different formulations of Agreement Statistics. All the statistical procedures utilized in this study attempted to represent the diagnostic agreement of clinicians on cases of reading disability. The Proportional Agreement statistic generally described the clues and diagnostic categories most frequently agreed upon by clinicians across all sessions conducted on the same case. The commonality statistics described the degree to which an individual clinician agreed with the cue selection or diagnostic judgments made by a group of clinicians across all sessions conducted on the same case. The Inter-Clinician Correlation statistic described the degree to which an individual clinician agreed with another clinician on the same case. The Intra-Clinician Correlation statistic described the degree to which a clinician agreed with him or her self across two sessions on the same case.

The basic research question was to describe the agreement in diagnostic judgments on the two instruments, SIMCASES and PAPERCASES, in order to determine the similarity or difference in clinician performance on these two instruments which are both simulated cases of reading disability but are different in format, or physical

construction, and in size, or amount of information included in the case.

CHAPTER IV

ANALYSIS OF THE DATA

Chapter IV analyzes the Clinicians' diagnostic problem-solving behavior on SIMCASES and PAPERCASES and answers the research question: Was clinical performance effected by the instrumentation, the form in which the case of reading disability was presented to the clinicians, SIMCASE or PAPERCASE. Performance on the two instruments will be reported within four sections of this chapter. Each of the four sections addresses the research question in terms of one measure of statistical agreement: Proportional Agreement, Commonality, Inter-Clinician Agreement or Intra-Clinician Agreement.

Proportional Agreement Question: Will the cues or diagnostic categories produced by the group of clinicians diagnosing a case of reading disability presented in the form of SIMCASES and PAPERCASES proportionally agree any more or less with the cues of diagnostic categories produced by previous groups of clinicians diagnosing a case of reading disability presented exclusively in the form of SIMCASES?

Commonality Question: Will the cues or diagnostic categories produced by an individual clinician diagnosing a case of reading disability presented in the form of a PAPERCASE agree any more or less with the cues or diagnostic categories produced by the group of clinicians diagnosing the same case of reading disability presented in the form of a SIMCASE?

Inter-Clinician Agreement Question: Will the Inter-Clinician Correlations on cues or diagnostic categories produced by comparing one clinician's session on the SIMCASE with another clinician's session on the SIMCASE be meaningfully different from the Inter-Clinician Correlations on cues or diagnostic categories produced by comparing one clinician's session on a SIMCASE with another clinician's session on the PAPERCASE form of the same case?

Intra-Clinician Agreement Question: Will the Intra-Clinician Correlations on cues or diagnostic categories produced by comparing a clinician with him/her self on one SIMCASE session and another SIMCASE session be meaningfully different from the Intra-Clinician Correlations on cues or diagnostic categories produced by comparing a clinician with him/her self on one SIMCASE session and a PAPERCASE session of the same case?

Proportional Agreement

The first column in Table 2 lists the diagnostic judgments mentioned in at least 25 percent of the observational sessions on Case I, or mentioned by clinicians in at least two of the eight sessions on Case I. The second column lists the frequency of a diagnostic judgment in terms of percentages and the third column translates the percentages into number of sessions. For example, "Language General - Strength" was mentioned 25 percent of the time by the group diagnosing this case or this category was identified as a strength in two of the eight observational sessions conducted on this case (see Appendix C for the standardized Diagnostic Checklist).

Table 2 specifies that 12 diagnostic categories (2, 3, 4, 6, 7, 7, 13, 14, 15, 19, 20, 21) were each identified by the groups diagnosing Case I in 25 percent of the observational sessions and five diagnostic categories (5, 10, 11, 16, 22) were identified in approximately 37 percent of the sessions. These percentages expressed insignificant Proportional Agreement on a collection of dispersive categories.

However, four of the remaining five diagnostic categories (9, 12, 17, 18) which were identified in at least 50 percent of the sessions on Case I suggested a low concentration of Proportional Agreement on a few

Table 2

The Proportional Agreement Statistic: Case I
Most Frequently Occurring Diagnostic Categories
in 8 Sessions among 4 Clinicians

Case: I	Diagnostic Category	% Sessions	No. Sessions
1	Attitude Toward School - Observation	.500	4
2	Attitude Toward Reading Instructional - Weakness	.250	2
3	Attitude Toward Reading Independent - Observation	.250	2
4	Language General - Strength	.250	2
5	Intellectual Potential General - Strength	.375	3
6	Oral Reading Rate - Weakness	.250	2
7	Oral Reading Fluency - Weakness	.250	2
8	Silent Reading Comprehension - Strength	.250	2
9	Word Analysis General - Weakness	.750	6
10	Phonetic Analysis General - Weakness	.375	3
11	Use of Letter-Sound Association - Weakness	.375	3
12	Use of Blends General - Weakness	.625	5
13	Use of Word Families - Weakness	.250	2
14	Use of Syllables - Weakness	.250	2
15	Ability to Blend Sounds into Whole Words - Weakness	.250	2
16	Word Recognition General - Weakness	.375	3
17	Word Recognition Basic Sight Words - Weakness	.500	4
18	Use of Initial Letters in Word Identification - Weakness	.500	4
19	Use of Medial Letters in Word Identification - Weakness	.250	2
20	Comprehension Oral - Strength	.250	2
21	Comprehension Listening - Strength	.250	2
22	Use of Context to Get Meaning - Strength	.375	3

diagnostic categories which at a very non-specific level touch upon the areas of reading difficulty represented in Case I. The group recorded Category 9: "Word Analysis General - Weakness" in five of the eight sessions. Dr. G. Sherman, an expert in reading diagnosis and a consultant with the Clinical Studies Project, informally diagnosed Case I and set forth Word Analysis as a problem area in Case I (personal communication). Likewise, Category 17: "Word Recognition Basic Sight Words - Weakness" and the related Category 18: "Use of Initial Letters in Word Identification - Weakness" were recorded in four of the eight sessions and Sherman also indicated Basic Sight Words as a reading problem area in Case I. Although Proportional Agreement was very infrequent, it appeared to be in a slightly meaningful direction, toward the general reading problem represented in this case.

Thus, while very few of the same diagnostic categories appeared across the eight sessions conducted on Case I, the four clinicians who diagnosed this case did agree at least 50 percent of the time on a few very broad categories which loosely connect with the actual areas of difficulty in this case.

The first column in Table 3 lists the diagnostic judgments mentioned in at least 25 percent of the observational sessions on Case II. The second column lists

Table 3

The Proportional Agreement Statistic: Case II
(Continued)

Case: II	Diagnostic Category	% Sessions	No. Sessions
22	Oral Reading Punctuation - Weakness	.250	2
23	Oral Reading Phrasing - Weakness	.250	2
24	Silent Reading Rate - Weakness	.375	3
25	Word Analysis General - Weakness	.375	3
26	Phonetic Analysis General - Strength	.250	2
27	Phonetic Analysis General - Weakness	.250	2
28	Use of Syllables - Weakness	.375	3
*29	Word Recognition General - Weakness	.750	6
30	Word Recognition General - Observation	.250	2
31	Word Recognition Basic Sight Words - Weakness	.375	3
*32	Use of Whole Word Approach - Weakness	.500	4
33	Use of Final Letters in Word Identification - Weakness	.250	2
34	Use of Letter Orders (Reversals) - Weakness	.375	3
35	Comprehension Oral - Strength	.250	2
36	Comprehension Silent - Observation	.250	2
37	Comprehension Vocabulary - Weakness	.250	2
38	Recall of Information - Observation	.250	2

Table 3

The Proportional Agreement Statistic: Case II
Most Frequently Occurring Diagnostic Categories
in 8 Sessions among 4 Clinicians

Case: II	Diagnostic Category	% Sessions	No. Sessions
1	Vision General - Strength	.250	2
2	Hearing General - Strength	.250	2
3	Attitude Toward Reading Independent - Weakness	.375	3
4	Home Environment - Strength	.250	2
5	Auditory Discrimination - Strength	.250	2
* 6	Visual Memory - Weakness	.500	4
7	Visual Discrimination - Strength	.250	2
* 8	Visual Discrimination - Weakness	.500	4
9	Grade Level Placement - Observation	.250	2
10	Intellectual Potential General - Strength	.375	3
11	Intellectual Potential General - Observation	.375	3
12	Intellectual Potential Verbal - Strength	.375	3
*13	Word Analysis General - Weakness	.500	4
14	Use of Suffixes - Weakness	.375	3
15	Use of Prefixes - Weakness	.375	3
16	Use of Syllables - Weakness	.375	3
17	Use of Syllables - Observation	.250	2
18	Oral Reading Rate - Weakness	.375	3
19	Oral Reading Rate - Observation	.250	2
*20	Oral Reading Fluency - Weakness	.625	5
21	Oral Reading Fluency - Observation	.375	3

the frequency of a diagnostic judgment in terms of percentages and the third column translates the percentages into number of sessions.

Table 3 specifies that 18 different diagnostic categories (1, 2, 4, 5, 7, 9, 17, 22, 23, 26, 27, 30, 33, 35, 36, 37, 38) were identified by the group diagnosing Case II in 25 percent of the observational sessions and 14 diagnostic categories (3, 10, 11, 12, 14, 15, 16, 18, 21, 24, 25, 28, 31, 34) were identified in approximately 37 percent of the sessions. As indicated in Table 2 of Case I, these percentages expressed insignificant Proportional Agreement on a collection of dispersive categories.

Five of the remaining six categories mentioned in at least 50 percent of the sessions again suggested a low concentration of Proportional Agreement on a few diagnostic categories which at a very non-specific level touch upon areas of reading difficulty represented in Case II. In particular, the group recorded Category 6: "Visual Memory - Weakness" and Category 8: "Visual Discrimination - Weakness" in four of the eight sessions. According to Sherman's informal diagnosis (personal communication), Case II included a problem in visual scanning and memory for word forms. A problem in the area of Category 20: "Oral Reading Fluency - Weakness," recorded in five of the eight sessions, was also

identified by Sherman. Furthermore, Category 13: "Word Analysis General - Weakness," as well as the most frequently mentioned Category 29: "Word Recognition General - Weakness" are general problem areas which Sherman elaborated upon in Case II.

Thus, while very few of the same diagnostic categories appeared across the eight sessions conducted on Case II, the four clinicians who diagnosed this case did agree at least 50 percent of the time on a few very broad categories which loosely connect with the actual areas of reading difficulty in Case II. Again, at best, the Proportional Agreement is very low and non-specific but in a slightly meaningful direction.

The first column in Table 4 lists the diagnostic judgments mentioned in at least 25 percent of the observational sessions on Case III. The second column lists the frequency of diagnostic judgment in terms of percentages and the third column translates the percentages into number of sessions.

Table 4 specifies that 18 diagnostic categories (3, 5, 7, 8, 9, 10, 11, 14, 20, 21, 24, 27, 29, 30, 31, 33, 34, 35) were identified by the group diagnosing Case III in 25 percent of the observational sessions and 11 diagnostic categories (1, 2, 6, 12, 15, 16, 18, 19, 23, 26, 32) were each identified in approximately 37 percent of the sessions. In other terms, as in

Table 4

The Proportional Agreement Statistic: Case III
Most Frequently Occurring Diagnostic Categories
in 8 Sessions among 4 Clinicians

Case: III	Diagnostic Categories	% Sessions	No. Sessions
1	Vision Acuity - Weakness	.375	3
2	Speech Articulation - Weakness	.375	3
3	Speech Articulation - Observation	.250	2
* 4	Motor Coordination - Weakness	.500	4
5	Attitude Toward School - Strength	.250	2
6	Relationship With Peers - Weakness	.375	3
7	Motivation for Reading - Observation	.250	2
8	Emotional Adjustment - Weakness	.250	2
9	Emotional Adjustment - Observation	.250	2
10	Parent-Child Relationship - Observation	.250	2
11	Grade Level Placement - Observation	.250	2
12	Use of Root Words - Weakness	.375	3
*13	Use of Syllables - Weakness	.500	4
14	Use of Syllables - Observation	.250	2
15	Oral Reading General - Weakness	.375	3
16	Oral Reading General - Observation	.375	3
*17	Oral Reading Fluency - Weakness	.500	4
18	Oral Reading Fluency - Observation	.375	3
19	Oral Reading Punctuation - Weakness	.375	3
20	Oral Reading Phrasing - Weakness	.250	2
21	Word Analysis General - Strength	.250	2
*22	Phonetic Analysis General - Strength	.500	4

Table 4

The Proportional Agreement Statistic: Case III
(Continued)

Case: III	Diagnostic Categories	% Sessions	No. Sessions
23	Phonetic Analysis General - Weakness	.375	3
24	Use of Letter-Sound Association - Strength	.250	2
*25	Use of Blends General - Weakness	.750	6
26	Word Recognition General - Strength	.375	3
27	Word Recognition General - Weakness	.250	2
*28	Word Recognition General - Observation	.500	4
29	Word Recognition Basic Sight Words - Weakness	.250	2
30	Word Recognition Basic Sight Words - Observation	.250	2
31	Comprehension General - Strength	.250	2
32	Comprehension General - Weakness	.375	3
33	Comprehension Oral - Strength	.250	2
34	Comprehension Listening - Strength	.250	2
35	Comprehension Vocabulary - Strength	.250	2

Case I and Case II, these percentages expressed an insignificant scatter of group agreement on a collection of categories.

Of the remaining six categories (4, 13, 17, 22, 25, 28) which were identified in at least 50 percent of the sessions, Category 13: "Use of Syllables - Weakness" and Category 25: "Use of Blends General - Weakness" are particularly relevant to Case III (Sherman, personal communication). These two categories which appeared in four of the eight sessions conducted on Case III suggest some Proportional Agreement but it appears to be only on a chance basis. In half of the sessions these categories were included in the diagnosis; in half of the sessions they were omitted from the diagnosis.

Table 5 summarizes the diagnostic categories which were included in at least 50 percent of the sessions conducted on Case I, Case II, and Case III. Across all three cases, the Proportional Agreement was very low. Very few diagnostic categories were repeatedly recorded by clinicians during the eight sessions conducted on each of the three cases of reading disability.

The five or six categories recorded most frequently in all cases were statements of "weakness" as opposed to statements of "strength" or "observations." When the clinician diagnosed a case, half of the time they mentioned five or six general diagnostic statements

Table 5

The Proportional Agreement Statistic
Diagnostic Categories of Highest Frequency
Across Cases: I, II, and III

Case I	%	No	Case II	%	No	Case III	%	No
Attitude Toward School - O	.500	4	Visual Memory - W	.500	4	Motor Coordination - W	.500	4
Word Analysis General - W	.750	6	Visual Discrimination - W	.500	4	Use of Syllables - W	.500	4
Use of Blends General - W	.625	5	Word Analysis General - W	.500	4	Oral Reading Fluency - W	.500	4
Word Recognition Basic Sight Words - W	.500	4	Oral Reading Fluency - W	.625	5	Phonetic Analysis General - S	.500	4
Use of Initial Letters in Word Identification - W	.500	4	Word Recognition General - W	.750	6	Use of Blends General - W	.750	6
			Use of Whole Word Approach - W	.500	4	Word Recognition General - O	.500	4

of weakness which are loosely connected with the actual reading disability in each case and half of the time the clinicians did not mention these same statements of weakness. The judgments mentioned most frequently were, at best, in a slightly meaningful direction; but, they were very few in number and very inconsistent in occurrence.

Table 6 reports the average amount of Proportional Agreement on diagnostic categories for each case. The mean for Case I, II, and III almost coincide. For all practical purposes, the mean proportion of the diagnosis agreed upon by the group for any case was very low, roughly .20.

The first column in Table 7 lists the cues selected by clinicians at least 50 percent of the observational sessions on Case I. The second column lists the frequency of cue selection in terms of percentages and the third column translates the percentages into number of sessions.

Twenty-three cues were most frequently collected from a cue inventory averaging 84 cues (see Appendix B for the SIMCASE Cue List and Appendix E for the PAPERCASE Cue List on Case I). Among the 23 cues selected in 50 percent of the sessions, 15 of the cues provided information in the form of a record or test booklet. The group of clinicians diagnosing Case I most frequently

Table 6

The Proportional Agreement Statistic
Average Proportional Agreement
On Diagnostic Categories
Across All Sessions per Case

Case	No. Sessions	Total No. DX Categories	Mean Proportional Agreement
Case I	8	75	.187
Case II	8	100	.204
Case III	8	96	.199

Table 7

The Proportional Agreement Statistic: Case I
Most Frequently Occurring Cue Categories
in 8 Sessions among 4 Clinicians

Case: I	Cue Category	% Sessions	No. Sessions
1	Background: Parent Form - Record Booklet	.875	7
2	Background: Teacher Form - Record Booklet	.625	5
3	Background: School Record - Record Booklet	.625	5
4	Background: School Information - Record Booklet	.500	4
5	Durrell Oral Reading: Examiner's Comments	.500	4
6	Durrell Oral Reading: Test Booklet	.875	7
7	Durrell Silent Reading: Test Booklet	.500	4
8	Durrell Word Recognition & Word Analysis: Test Booklet	.500	4
9	Durrell Hearing Sounds in Words-Primary: Test Booklet	.500	4
10	Durrell Phonetic Spelling of Words: Test Booklet	.500	4
11	Ekwall Phonics Survey Modified: Examiner's Comments	.500	4
12	Ekwall Phonics Survey Modified: Test Booklet	.750	6
13	Gates McKillop Auditory Blending: Test Booklet	.500	4
14	Attitude: Sentence Completion - Test Booklet	.500	4
15	Slosson Oral Reading Test: Test Scores	.625	5
16	Slosson Oral Reading Test: Test Booklet	.625	5
17	WISC Verbal Scale: Test Scores	.750	6
18	WISC Verbal Scale: Examiner's Comments	.500	4
19	WISC Full Scale: Test Scores	.625	5
20	Audiometric Record: Examiner's Comments	.625	5
21	Audiometric Record: Test Booklet	.500	4
22	Vision Test: Examiner's Comments	.625	5
23	Vision Test: Test Booklet	.500	4

chose to examine information in the form of a record or test booklets which described student behavior as opposed to referring to other forms of information such as test scores and examiner's comments. To the very limited degree of 15 cues selected 50 percent of the time, clinicians relied more upon a qualitative description of reading behavior rather than a quantitative score or the post hoc comments provided by the examiner who administered the texts.

The first four (1, 2, 3, 4) of the 23 most frequent cues selected on Case I referred to background information, four other cues (5, 6, 15, 16) referred to oral reading information, and at least an additional six cues (8, 9, 10, 11, 12, 13) referred to information on word analysis. Cue selection seemed to cluster in three areas of information: background, oral reading, and word analysis.

In all, the 23 cues selected in 50 percent of the sessions suggested a low concentration of Proportional Cue Agreement for Case I which exceeded the Proportional Diagnostic Agreement previously reported for Case I. Only five diagnostic categories were agreed upon in 50 percent of the sessions while 23 cue categories were agreed upon in 50 percent of the sessions. The cues collected were more frequently the same than

the diagnostic judgments recorded about Case I. However, although agreement was relatively higher for cues than for diagnostic categories, the overall degree of Proportional Agreement was quite low, a 50-50 chance occurrence of a few of the same categories.

The first column in Table 8 lists the cues selected by clinicians in at least 50 percent of the sessions on Case II. The second column lists the frequency of cue selection in terms of percentages and the third column translates the percentages into number of sessions. Nineteen cues were most frequently collected from a cue inventory averaging 73 cues (see Appendix B for the SIMCASE Cue List and Appendix E for the PAPERCASE Cue List on Case II).

As previously observed in Case I, the type of information most frequently referred to during the sessions conducted on Case II was the record or test booklet. The group of clinicians diagnosing Case II most frequently chose to examine the record or test booklet which described student behavior as opposed to referring to other forms of information such as test scores and examiner's comments. It seems reasonable to suggest again, to the limited degree of 11 cues selected 50 percent of the time, this group also appeared to be relying somewhat more upon a qualitative description of reading behavior rather than

Table 8

The Proportional Agreement Statistic: Case II
Most Frequently Occurring Cue Categories
in 8 Sessions among 4 Clinicians

Case: II	Cue Categories	% Sessions	No. Sessions
1	Background: Biographical Data - Record Booklet	.625	5
2	Background: Physical/Health - Record Booklet	.875	7
3	Background: Home/Family - Record Booklet	.750	6
4	Background: Classroom Information - Record Booklet	.870	7
5	Dolch Word List: Test Booklet	.750	6
6	Durrell Oral Reading: Test Booklet	.750	6
7	Durrell Silent Reading: Test Scores	.500	4
8	Durrell Silent Reading: Examiner's Comments	.500	4
9	Durrell Silent Reading: Test Booklet	.500	4
10	Durrell Word Recognition & Word Analysis: Test Booklet	.625	5
11	Durrell Visual Memory of Words-Primary: Test Booklet	.625	5
12	Gates-MacGinitie Comprehension: TestScores	.500	4
13	Gates McKillop Recognizing & Blending Common Word Parts: Test Scores	.500	4
14	Gates McKillop Recognizing & Blending Common Word Parts: Test Booklet	.750	6
15	Informal Oral Reading: Examiner's Comments	.500	4
16	Slosson Oral Reading: Examiner's Comments	.500	4
17	Slosson Oral Reading: Test Booklet	.750	6
18	WISC Full Scale: Test Scores	.500	4
19	WISC Verbal Scale: Test Scores	.500	4

a quantitative score or the post hoc comments provided by the examiner who administered the tests.

Also, resembling Case I, the cues selected on Case II seemed to cluster in two areas. The first four (1, 2, 3, 4) of the 19 most frequently collected cues referred to background information and four other cues (6, 15, 16, 17) referred to oral reading information.

The 19 cues selected in 50 percent of the sessions suggested a low concentration of Proportional Cue Agreement for Case II which exceeded the Proportional Diagnostic Agreement previously reported for Case II. Only six diagnostic categories were agreed upon in 50 percent of the sessions while 19 cue categories were agreed upon in 50 percent of the sessions. The same information was collected more often than the same diagnostic statements were written about Case II. In general, there was somewhat more agreement within the group on what type of information to collect than on how to describe the reading disability presented in the case. Although Proportional Agreement was relatively higher for cues than for diagnostic categories, the amount of Proportional Agreement remains very low.

The first column in Table 9 lists the cues selected by clinicians in at least 50 percent of the sessions on Case III. The second column lists the

Table 9

The Proportional Agreement Statistic: Case III
Most Frequently Occurring Cue Categories
in 8 Sessions among 4 Clinicians

Case: III	Cue Categories	% Sessions	No. Sessions
1	Background: Parent Form - Examiner's Comments	.875	7
2	Background: Teacher Form - Examiner's Comments	.750	6
3	Background: School Record - Examiner's Comments	.500	4
4	Background: Pupil Progress Report - Examiner's Comments	.500	4
5	Durrell Oral Reading: Test Scores	.500	4
6	Durrell Oral Reading: Test Booklet	.750	6
7	Durrell Listening Comprehension: Test Scores	.625	5
8	Gates McGinitie Speed & Accuracy - Test Scores	.500	4
9	Gates McKillop Recognizing & Blending Common Word Parts: Test Scores	.500	4
10	Gates McKillop Recognizing & Blending Common Word Parts: Test Booklet	.625	5
11	Gates McKillop Nonsense Words: Test Booklet	.500	4
12	Slosson Oral Reading Test: Test Scores	.500	4
13	Slosson Oral Reading Test: Test Booklet	.500	4

frequency of cue selection in terms of percentages and the third column translates the percentages into number of sessions. Thirteen cues were most frequently collected from the cue inventory averaging 91 cues (see Appendix B for the SIMCASE Cue List and Appendix E for the PAPERCASE Cue List on Case III).

Contrary to Case I and Case II, the group of clinicians diagnosing Case III did not choose to examine record of test booklets more often than other forms of information. Instead, cues were almost equally distributed among three forms of information: test booklets, test scores, and examiner's comments. However, like Case I and Case II, four (1, 2, 3, 4) of the 13 most frequently collected cues referred to background information and four other cues (5, 6, 12, 13) referred to oral reading information. No other clusters of cues were observed.

The 13 cues selected in 50 percent of the sessions indicated very low Proportional Cue Agreement for Case III. The amount of cue agreement observed in Case III exceeded Proportional Diagnostic Agreement by only seven categories. Thirteen cue categories were agreed upon in 50 percent of the sessions and six diagnostic categories were agreed upon in 50 percent of the sessions. This observation was contrary to Case I and

Case II in which the number of most frequent cue categories exceeded the number of most frequent diagnostic categories by a wider margin.

Table 10 reports the average amount of Proportional Agreement on cue categories for all three cases. The mean proportion of cues agreed upon by the group across all three cases was approximately .30. The average frequency with which the same cues were selected on a case was only slightly higher than the average frequency with which the same diagnostic judgments were recorded on a case, approximately .20 (see Table 6).

Summary of Proportional Agreement Results

The case by case examination of the diagnostic categories most frequently agreed upon by the groups uncovered percentages of group agreement which were constantly low. The average amount of Proportional Agreement per case was approximately 20 percent. For each case, only five of six of the same diagnostic categories appeared in 50 percent of the sessions. Half of the time the clinicians included five or six of the same judgments in their diagnosis and half of the time these same judgments did not appear in their diagnosis. For example, in four sessions the clinicians identified "Word Analysis" as a weakness in a case and in four other sessions they did not mention "Word Analysis" as a

Table 10

The Proportional Agreement Statistic
Average Proportional Agreement
on Cue Categories
Across All Sessions Per Case

Case	No. Sessions	Total No. CX Categories	Mean Proportional Agreement
Case I	8	72	.323
Case II	8	58	.391
Case III	8	70	.288

weakness. There was only a 50-50 chance agreement that "Word Analysis" was a weakness, far from a group consensus on a possible area of difficulty in the case. The five or six most frequently mentioned diagnostic categories were mostly general statements of weakness which were loosely connected with the actual areas of reading difficulty in each case.

The case by case examination of the cue categories most frequently agreed upon by the group uncovered percentages of group agreement on cues to collect which slightly exceeded group agreement on diagnostic judgments. The average Proportional Agreement per case on cues was approximately .30. For each case, from 13 to 23 cue categories, as opposed to only five or six diagnostic categories, appeared in 50 percent of the sessions. In Case I and Case II clinicians most frequently collected information in the form of record or test booklets. In all three cases, there was a low concentration of cues in the areas of background information and oral reading behavior. Four different cues on background information and four different cues on oral reading were collected in 50 percent of the sessions per case.

In sum, the very low Proportional Agreement on diagnostic and cue categories in this study bears a strikingly close resemblance to the very low Proportional

Agreement on diagnostic and cue categories found in the original Clinical Studies Project observational study, OS77.1.

Commonality Score

Table 11 reports the average Commonality Scores on diagnostic categories for each case. This statistic described the degree to which any individual session on a case included the diagnostic categories most frequently recorded during all other sessions on a case. The mean Commonality Scores displayed in the second column were derived by comparing the diagnostic categories within each individual session on a case with the most frequent diagnostic categories in all other sessions on a case, irrespective of the observational instrument. The mean Commonality Scores in column three were derived by comparing the diagnostic categories within sessions utilizing the SIMCASE instrumentation with the most frequent categories in all other sessions on a case. The mean Commonality Scores in column four were derived by comparing the diagnostic categories within sessions utilizing the PAPERCASE instrumentation with the most frequent categories in all other sessions on a case.

The second column of Table 11 indicates that, on the average, any individual session included approximately 57 percent or half of the diagnostic judgments

Table 11

The Commonality Statistic: Cases I, II and III
Individual vs. Group Agreement on Diagnostic Categories
in 8 Sessions among 4 Clinicians

Case	All Sessions	SIMCASE Sessions	PAPER CASE Sessions
I	$\bar{X}$ = .567%	$\bar{X}$ = .494%	$\bar{X}$ = .785%
II	$\bar{X}$ = .574%	$\bar{X}$ = .555%	$\bar{X}$ = .632%
III	$\bar{X}$ = .571%	$\bar{X}$ = .519%	$\bar{X}$ = .727%

which were most frequently recorded for the same case during all other sessions. In other terms, each session on a case had 57 percent of the most frequently recorded diagnostic judgments in common with all other sessions on the same case. It is interesting to observe that the Commonality Scores for sessions exclusively utilizing the PAPERCASES were slightly above the mean Commonality Score for sessions exclusively utilizing the SIMCASES. On the average, individual sessions on PAPERCASES included from 63 to 78 percent of the most frequently recorded diagnostic judgments in common with all other sessions on a case while individual sessions on SIMCASES included from 49 to 55 percent of the most frequently recorded diagnostic judgments in common with all other sessions on a case.

Table 12 reports the average Commonality Scores on cue categories for each case. This statistic described the degree to which any individual session on a case included the cue categories most frequently recorded during all other sessions on a case. The mean Commonality Scores displayed in the second column were derived by comparing the cue categories within each individual session on a case with the most frequent cue categories in all other sessions on a case, irrespective of the observational instrument. The mean Commonality Scores in column three were derived by comparing the cue categories

Table 12

The Commonality Statistic: Case I, II, and III
Individual vs. Group Agreement on Cue Categories
in 8 Sessions among 4 Clinicians

Case	All Sessions SIMCASE & PAPER CASE	SIMCASE Sessions Only	PAPER CASE Sessions only
I	$\bar{X}$ = .694%	$\bar{X}$ = .706%	$\bar{X}$ = .657%
II	$\bar{X}$ = .691%	$\bar{X}$ = .658%	$\bar{X}$ = .787%
III	$\bar{X}$ = .620%	$\bar{X}$ = .599%	$\bar{X}$ = .680%

within sessions utilizing the SIMCASE instrumentation with the most frequent categories for all other sessions on a case. The mean Commonality Scores in column four were derived by comparing the cue categories within sessions utilizing the PAPERCASE instrumentation with the most frequent categories in all other sessions on a case.

The second column of Table 12 indicates that, on the average, an individual session included from 62 percent to 69 percent of the cues which were most frequently recorded for the same case during all other sessions. The similarity among the cue Commonality Scores for all sessions, for sessions utilizing SIMCASES and for sessions utilizing PAPERCASES is an important observation. The percentages are not meaningfully different regardless of the instrumentation.

Summary of Commonality Results

The case by case examination of the mean Commonality Scores led to the finding that individual sessions contained approximately 57 percent or half of the diagnostic categories which were most frequently recorded for the same case during all other sessions and from 62 percent to 69 percent of the cue categories which were most frequently recorded for the same case during all other sessions.

The mean diagnostic Commonality Scores for sessions utilizing the PAPERCASES were not meaningfully different from sessions utilizing the SIMCASES nor were the mean cue Commonality Scores for sessions utilizing the PAPERCASES meaningfully different from sessions utilizing the SIMCASES. For both diagnostic and cue categories the mean Commonality Scores derived from PAPERCASE sessions were similar to the scores derived from SIMCASE sessions. Diagnostic problem-solving performance on the two instruments was very similar in terms of the Commonality dimension of statistical agreement.

Inter-Clinician Correlational Agreement

Table 13 contains all the intra (4) and inter (24) phi correlation coefficients on diagnostic categories for Case I. The 28 correlations range from .00 to .23 with an overall mean Inter/Intra Correlation of .096 and a standard deviation of .069. The correlations between sessions were generally very low and signal only the slightest degree of relationship or statistical agreement on diagnostic categories in Case I.

Table 14 contains all the possible intra (4) and inter (24) phi correlation coefficients on diagnostic categories for Case II. The 28 correlations ranged from .01 to .26 with an overall mean Inter/Intra Correlation

Table 13

Inter- and Intra-Agreement Statistic:
Case I Correlation Matrix on Diagnostic Categories

	S 501(1)	S 501(2)	S 502(1)	S 502(2)	S 507(1)	P 507(2)	P 510(1)	S 510(2)
S 501(1)		.06	.04	.04	.05	.00	.00	.00
S 501(2)			.06	.11	.07	.16	.19	.03
S 502(1)				.15	.02	.07	.09	.00
S 502(2)					.13	.18	.23	.18
S 507(1)						.14	.12	.10
P 507(2)							.22	.08
P 510(1)								.17
S 510(2)								

501 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

Table 14

Inter- and Intra-Agreement Statistic:
Case II Correlation Matrix on Diagnostic Categories

	S 503(1)	S 503(2)	S 504(1)	S 504(2)	S 508(1)	P 508(2)	P 511(1)	S 511(2)
S 503(1)		.13	.08	.07	.06	.01	.08	.02
S 503(2)			.05	.04	.06	.07	.11	.04
S 504(1)				.26	.17	.05	.07	.03
S 504(2)					.14	.13	.17	.17
S 508(1)						.12	.19	.20
P 508(2)							.14	.14
P 511(1)								.21
S 511(2)								

503 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

of .10 and a standard deviation of .06. Once again, the correlations between sessions were very low, indicating only the slightest degree of statistical agreement on diagnostic categories in Case II.

Table 15 contains all the possible intra (4) and inter (24) phi correlation coefficients on diagnostic categories for Case III. The 28 correlations ranged from .02 to .26, very similar to the range for Case I and Case II. The mean Inter/Intra Correlation was .10 with a standard deviation of .06. As previously indicated for Case I and Case II, the low correlations between sessions provided very limited evidence of statistical agreement on diagnostic categories in Case III.

Table 16 contains all the intra (4) and inter (24) phi correlation coefficients on cue categories for Case I. The 28 correlations range from .12 to .48 with an overall mean Inter/Intra Correlation of .23 and a standard deviation of .07. The correlations between sessions on cue categories in Case I were somewhat higher than the correlations between sessions on diagnostic categories.

Table 17 contains all the intra (4) and inter (24) phi correlation coefficients on cue categories for Case II. The 28 correlations range from .08 to .61 with an overall mean Inter/Intra Correlation of .27 and a

Table 15

Inter- and Intra-Agreement Statistic:
Case III Correlation Matrix on Diagnostic Categories

	S 502(1)	S 505(2)	S 506(1)	S 506(2)	S 509(1)	P 509(2)	P 512(1)	S 512(2)
S 505(1)		.15	.02	.05	.05	.05	.07	.05
S 505(2)			.07	.13	.13	.08	.15	.07
S 506(1)				.02	.05	.10	.14	.12
S 506(2)					.02	.09	.08	.22
S 509(1)						.09	.07	.02
P 509(2)							.26	.12
P 512(1)								.13
S 512(2)								

505 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

Table 16

Inter- and Intra-Agreement Statistics:
Case I Correlation Matrix on Cue Categories

	S 501(1)	S 501(2)	S 502(1)	S 502(2)	S 507(1)	P 507(2)	P 510(1)	S 510(2)
S 501(1)		.20	.17	.21	.25	.14	.16	.26
S 501(2)			.26	.23	.25	.28	.28	.28
S 502(1)				.31	.26	.19	.14	.20
S 502(2)					.29	.19	.14	.30
S 507(1)						.36	.25	.19
P 507(2)							.48	.12
P 510(1)								.19
S 510(2)								

501 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

Table 17

Inter- and Intra-Agreement Statistics:
Case II Correlation Matrix on Cue Categories

	S 503(1)	S 503(2)	S 504(1)	S 504(2)	S 508(1)	P 508(2)	P 511(1)	S 511(2)
S 503(1)		.61	.33	.13	.08	.22	.11	.08
S 503(2)			.37	.19	.22	.30	.21	.13
S 504(1)				.43	.37	.29	.30	.16
S 504(2)					.33	.22	.29	.25
S 508(1)						.27	.33	.28
P 508(2)							.55	.24
P 511(1)								.34
S 511(2)								

503 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

standard deviation of .12. Similar to Case I, correlations between sessions on cue categories in Case II were somewhat higher than the correlations between sessions on diagnostic categories.

Table 18 contains all the intra (4) and inter (24) phi correlation coefficients on cue categories for Case III. The 28 correlations ranged from .02 to .43 with an overall mean Inter/Intra Correlation of .18 and, as in Cases I and II, the correlations between sessions on cue categories in Case III were slightly higher than the correlations between sessions on diagnostic categories.

Table 19 reports the average Inter-Agreement statistic on diagnostic categories for each case. This statistic summarizes the degree to which clinicians agreed with each other on diagnostic categories when diagnosing two slightly different versions of the same reading problem, original or replicate case, utilizing two observational instruments which were different in format and size, SIMCASE and/or PAPERCASE.

Column two of Table 19 presents the mean Inter-Agreement derived from the 13 correlations between a SIMCASE session on a case and another SIMCASE session on the same case. Column three presents the mean Inter-Agreement derived from the 10 correlations between a

Table 18

Inter- and Intra-Agreement Statistics:
Case III Correlation Matrix on Cue Categories

	S 505(1)	S 505(2)	S 506(1)	S 506(2)	S 509(1)	P 509(2)	P 512(1)	S 512(2)
S 505(1)		.36	.19	.02	.16	.27	.18	.14
S 505(2)			.22	.09	.27	.40	.15	.16
S 506(1)				.24	.21	.23	.06	.12
S 506(2)					.10	.05	.03	.03
S 509(1)						.39	.09	.13
P 509(2)							.26	.33
P 512(1)								.43
S 512(2)								

505 = Subject; (1) = Session Number; S = SIMCASE; P = PAPERCASE

Table 19

The Inter-Agreement Statistic
Inter-Mean Coefficients on Diagnostic Categories
Cases I, II, and III

Case	No.	SIMCASE Sessions Inter Ø = 13	No.	SIMCASE & PAPERCASE Sessions Inter Ø = 10	No.	PAPERCASE Sessions Inter Ø = 1	No.	All Sessions Inter Ø = 24
I		$\bar{X}$ = .06 σ = .05		$\bar{X}$ = .11 σ = .07		$\bar{X}$ = .22		$\bar{X}$ = .09 σ = .07
II		$\bar{X}$ = .08 σ = .06		$\bar{X}$ = .10 σ = .06		$\bar{X}$ = .14		$\bar{X}$ = .09 σ = .05
III		$\bar{X}$ = .07 σ = .05		$\bar{X}$ = .09 σ = .03		$\bar{X}$ = .26		$\bar{X}$ = .09 σ = .05

SIMCASE session on a case and a PAPERCASE session on the same case. Column four presents the mean Inter-Agreement derived from the correlation of one PAPERCASE session on a case with another PAPERCASE session on the same case. Column five presents the mean Inter-Agreement derived from all 24 correlations of sessions on a case irrespective of the observational instruments utilized in the sessions.

To begin with, for all sessions, regardless of instrumentation or the case, one clinician's agreement with another clinician's diagnostic judgments averaged .09 which is only slightly higher than zero agreement. Clinicians rarely agreed with each other on their diagnostic statements about a case of reading disability. Separating the sessions by instrumentation, the mean inter-correlation between one SIMCASE session and another SIMCASE session was not quantitatively different from the mean inter-correlations between a SIMCASE session and a PAPERCASE session. The correlations were at or close to .10.

The inter-correlations between one PAPERCASE and another PAPERCASE has the potential to be misleading since it does not represent an average value. The value reported was based on only one correlation per case. Referring back to the correlation matrixes on diagnostic

cues for each case (Tables 13, 14 and 15), there were inter-correlations between one SIMCASE session and another SIMCASE session which were at or around .26.

Table 20 reports the average Inter-Agreement statistic on cue categories for each case. This statistic summarizes the degree to which clinicians agreed with each other on the cues to select or types of information to exaimine when presented with two slightly differenc versions of the same reading problem, original or replicate case, utilizing two observational instruments which were different in format and size, SIMCASE and/or PAPERCASE.

Column two of Table 20 presents the mean Inter-Agreement derived from the 13 correlations of one SIMCASE session on a case with another SIMCASE session on the same case. Column three presents the mean Inter-Agreement derived from the 10 correlations of a SIMCASE session on a case with a PAPERCASE session on the same case. Column four presents the Inter-Agreement derived from the correlation of one PAPERCASE session on a case with another PAPERCASE session on the same case. Column five presents the mean Inter-Agreement derived from all 24 correlations of sessions on a case irrespective of the observational instrument utilized in the sessions.

Table 20

The Inter-Agreement Statistic
Inter-Mean Phi Coefficients on Cue Categories
Cases I, II, and III

Case	SIMCASE Sessions No. Inter Ø = 13	SIMCASE & PAPERCASE Sessions No. Inter Ø = 10	PAPERCASE Sessions No. Inter Ø = 1	All Sessions No. Inter Ø = 24
I	$\bar{X}$ = .24 σ = .03	$\bar{X}$ = .18 σ = .06	$\bar{X}$ = .48	$\bar{X}$ = .23 σ = .07
II	$\bar{X}$ = .22 σ = .10	$\bar{X}$ = .25 σ = .06	$\bar{X}$ = .55	$\bar{X}$ = .24 σ = .10
III	$\bar{X}$ = .14 σ = .07	$\bar{X}$ = .17 σ = .12	$\bar{X}$ = .26	$\bar{X}$ = .16 σ = .09

Considering all sessions, all three cases, and both forms of instrumentation, the average agreement in cue selection found between one clinical session was in a low positive range. Only to the limited degree expressed by the average inter-correlations of .23, .24, and .16 did clinicians agree with each other on information to examine in the process of diagnosing a case of reading disability.

Separating the sessions by instrumentation, the mean inter-correlations between one SIMCASE session and another SIMCASE session were not quantitatively different from the mean inter-correlations between a SIMCASE session and a PAPERCASE session. However, the inter-correlations on cue categories were closer to .20 while the inter-correlations on diagnostic categories were closer to .10. Hence, the average agreement between clinicians on cues just barely exceeded the average agreement between clinicians on diagnosis.

As observed under diagnostic categories, the inter-correlation reported between one PAPERCASE and another PAPERCASE were based on only one correlation per case. While these inter-correlations on cues are not averages they are noteworthy. Referring back to the data matrixes (Tables 16, 17 and 18), .48 is the highest inter-correlation obtained for Case I and .55

is the highest inter-correlation obtained for Case II. These values hinted at the possibility that clinicians may have selected more of the same cues when using the PAPERCASE instrumentation which allowed them to independently locate the information in a notebook. The format of the SIMCASE required the clinicians to request information from the experimenter who located it in a file box and handed it to the clinician.

Summary of Inter-Agreement Results

The case by case examination of the agreement between clinical sessions on diagnostic categories uncovered inter-correlation coefficients which were constantly in the low positive range. The average Inter-Agreement per case was .09, which is only slightly higher than zero agreement between sessions on diagnostic judgments about the case. Comparing performance on the two instruments, the mean inter-correlations for diagnostic categories between two SIMCASE sessions was not quantitatively different from the mean inter-correlations for diagnostic categories between a SIMCASE session and a PAPERCASE session. These inter-correlations were at or close to .10.

The case by case examination of the agreement between clinical sessions on cue categories uncovered

inter-correlations which slightly exceeded the inter-correlations on diagnostic categories. The inter-correlations on cue categories were closer to .20 while the inter-correlations on diagnostic categories were closer to .10. Comparing performance on the two instruments, the mean inter-correlations for cue categories between two SIMCASE sessions were not quantitatively different from the mean inter-correlations on cue categories between a SIMCASE session and a PAPERCASE session. Diagnostic problem-solving performance on the two instruments was very similar in terms of the Inter-Correlation dimension of statistical agreement.

Intra-Correlational Agreement

Table 21 reports the average Intra-Agreement statistic on diagnostic categories. The mean intra-correlations summarize the degree to which clinicians agreed with themselves when diagnosing two slightly different versions of the same reading problem, original or replicate case, utilizing observational instruments which were different in format and size, SIMCASE and/or PAPERCASE.

Column two of Table 21 presents the mean intra-correlation for the two clinicians who were assigned to the SIMCASE form of a case in one session and a SIMCASE form of the same case in another session. Column three

Table 21

The Intra-Agreement Statistic
Intra-Mean Phi Coefficients on Diagnostic Categories
Cases I, II, and III

Case	SIMCASE Sessions Intra $\emptyset$ = 2	SIMCASE & PAPER CASE Sessions Intra $\emptyset$ = 2	All Sessions Intra $\emptyset$ = 4
I	$\bar{X}$ = .10 σ = .06	$\bar{X}$ = .15 σ = .02	$\bar{X}$ = .13 σ = .04
II	$\bar{X}$ = .19 σ = .09	$\bar{X}$ = .16 σ = .06	$\bar{X}$ = .18 σ = .06
III	$\bar{X}$ = .08 σ = .09	$\bar{X}$ = .11 σ = .02	$\bar{X}$ = .09 σ = .05

reports the mean intra-correlation for the two clinicians who were assigned to a SIMCASE form of a case in one session and a PAPERCASE form of the same case in another session. Column four reports the overall mean intra-correlation for the four clinicians assigned to a case regardless of the instruments utilized during the sessions.

Comparing clinicians with themselves on diagnostic judgments, the average Intra-Agreement over all four sessions was .13 for Case I, .18 for Case II, and .09 for Case III. These mean intra-correlations on diagnostic categories were similar to the mean inter-correlation on diagnostic categories which was .09 (see Table 19, p. 91). The average Inter- and Intra-Agreement for all three cases was remarkably alike. The clinicians did not agree with themselves (Intra) on diagnostic judgments any more than they agreed with each other (Inter) on diagnostic judgments.

Separating the sessions by instrumentation, the mean intra-correlation between one SIMCASE session and another SIMCASE session were not quantitatively different from the mean intra-correlations between a SIMCASE session and a PAPERCASE session. Without exception, the correlations remained very low, and nearly the same.

Table 22 reports the average Intra-Agreement statistic on cue categories. The intra-correlations summarize the degree to which clinicians agreed with themselves on the cues to select or types of information to examine when presented with two slightly different versions of the same reading problem, original or replicate case, utilizing observational instruments which were different in format and size, SIMCASE and/or PAPERCASE.

Column two of Table 22 reports the mean intra-correlation for the two clinicians who were assigned to the SIMCASE form of a case in one session and a SIMCASE form of the same case in another session. Column three reports the mean intra-correlations for the two clinicians who were assigned to a SIMCASE form of a case in one session and a PAPERCASE form of the same case in another session. Column four reports the overall mean intra-correlation for the four clinicians assigned to a case regardless of the instruments utilized during the sessions.

Comparing clinicians with themselves on cue selections, the average agreement over all sessions was .26 for Case I, .41 for Case II, and .35 for Case III. These mean inter-correlations are higher than the mean inter-correlations on diagnostic categories which were .13 for Case I, .18 for Case II, and .09 for

Table 22

The Intra-Agreement Statistic
Intra-Mean Phi Coefficients on Cue Categories
Cases I, II, and III

Case	SIMCASE Sessions Intra $\emptyset = 2$	SIMCASE & PAPER CASE Sessions Intra $\emptyset = 2$	All Sessions Intra $\emptyset = 4$
I	$\bar{X} = .25$ $\sigma = .07$	$\bar{X} = .27$ $\sigma = .12$	$\bar{X} = .26$ $\sigma = .08$
II	$\bar{X} = .52$ $\sigma = .12$	$\bar{X} = .30$ $\sigma = .04$	$\bar{X} = .41$ $\sigma = .14$
III	$\bar{X} = .30$ $\sigma = .08$	$\bar{X} = .41$ $\sigma = .02$	$\bar{X} = .35$ $\sigma = .08$

Case III. These data suggest that clinicians tended to agree a little more often with themselves than with each other on cues to examine about a case of reading disability. Clinicians agreed most with themselves on the cues to Select in Case II.

Separating the sessions by instrumentation, the mean intra-correlations between one SIMCASE session and another SIMCASE session were generally not quantitatively different from the mean intra-correlations between a SIMCASE session and a PAPERCASE session. However, in Case II, the correlation of .52 between SIMCASE sessions appears to be significantly higher than the correlation of .30 between a SIMCASE and PAPERCASE session.

Summary of Intra-Agreement Results

The case by case examination of clinician's agreement with themselves on diagnostic categories uncovered intra-correlation coefficients which were constantly in the low positive range. The average Intra-Agreement was .13 for Case I, 18 for Case II, and .09 for Case III. On cue categories, the Intra-Agreement was higher for each case. The mean intra-correlation on cues for Case I was .26, Case II was .41, and Case III was .35.

The mean intra-correlations for diagnostic or cue categories between two SIMCASE sessions were not quantitatively different from the mean intra-correlations for diagnostic or cue categories between a SIMCASE session and a PAPERCASE session. Diagnostic problem-solving performance on the two instruments was very similar in terms of the Intra-Correlation dimension of statistical agreement.

CHAPTER V

CONCLUSIONS AND IMPLICATIONS

This study was an extension of the research initiated by the Clinical Studies Project within the Institute for Research on Teaching (IRT). It sought additional information of value in understanding the diagnostic problem-solving performance of reading clinicians and pursued another research question: "Whether the group administered PAPERCASES could be substituted for the less efficient individually administered SIM-CASES as an instrument in collecting data on diagnostic problem-solving performance?" The answer to this question was based upon a comparison of the clinicians' performance on the two instruments. The criteria for comparison was statistical agreement data.

This chapter presents the major conclusions resulting from the statistical agreement data and calls attention to some implications of these conclusions for future research and for training reading specialists in the clinical process of diagnosing cases of reading disability.

Conclusions

Format of PAPERCASES and SIMCASES

In terms of the Inquiry Theory of Clinical Problem Solving, the form of simulation exercise did not affect the outcome of the Clinical Encounter. According to Vinsonhaler, et al. (1977), the Clinical Encounter includes: (1) the Clinical Case, (2) the Clinician, and (3) the Clinical Interaction. The clinicians' interaction with the clinical case in the form of a PAPERCASE or a SIMCASE did not affect the statistical agreement, the criteria for comparing performance on the instruments. The more efficient PAPERCASES, which permitted group administration, generated statistical agreement data which was very similar to that derived from the SIMCASES which required individual administration.

After the sessions were completed, some clinicians said they felt more at ease working on their own in the PAPERCASE booklet while others said they enjoyed having the information handed to them from the SIMCASE file box. However, the clinicians did not feel that the form in which the case was presented to them made the case any easier or more difficult to diagnose. In fact, the data show that clinicians did not agree any more or less with themselves or with each other when utilizing either form of simulation exercise. The PAPERCASE

proved to be more efficient and appeared to be at least as sensitive an instrument for observing diagnostic problem-solving performance.

Size or Cue Content of PAPERCASES and SIMCASES

The PAPERCASES contained four different categories of information while the SIMCASES contained six different categories of information about each case of reading disability. Although the PAPERCASES contained roughly 20 percent fewer cue categories than the SIMCASES, the inter and intra cue or diagnostic agreement between a PAPERCASE session and a SIMCASE session was not quantitatively different from the inter and intra cue or diagnostic agreement between two SIMCASE sessions on the same case. Less information did not make a quantitative difference in the correlational agreement data on each case. There appeared to be one exception to this generalization; the intra-correlation on cues for clinical subject 503 which was .61. This remarkably higher intra-correlation is more likely a descriptor of an individual's consistency in approach to a case than any indication of the usefulness of the additional cues contained in the SIMCASE.

Since both SIMCASES and PAPERCASES had a comprehensive data base including all the cues judged necessary to diagnose the reading disability presented in a case,

the additional cues in the SIMCASE were essentially redundant forms of information. In all, there was more information contained in either instruments than any clinician chose to examine during the standardized thirty minute data collection time period.

Statistical Agreement on PAPERCASES and SIMCASES

While the instruments were different in format and size or cue content, the performance of clinicians on both instruments resulted in consistently low positive estimates of statistical agreement. The inter- and intra-correlations on cues and diagnostic categories suggests that the two instruments are parallel forms in the sense that they yeilded similar correlation coefficients, with the one possible exception of a mean intra cue correlation between two SIMCASE sessions on Case II ($\bar{X}$ = .52) in comparison with the correlation between a PAPERCASE and a SIMCASE ($\bar{X}$ = .30). Generally, the same problem-solving behavior was elicited from the two instruments. The PAPERCASES proved to be an equally as effective instrument for observing diagnostic problem-solving behavior.

Statistical Agreement on Diagnostic Categories

On all three cases, clinicians did not agree any more with themselves than with each other on diagnostic

judgments. The intra- and inter-correlation coefficients essentially ranged from zero to low positive. While the correlations were not quantitatively different, they were stable and qualitatively meaningful. They described very inconsistent performance in diagnosing cases of reading disability. Although the cues remained the same, the clinicians constantly changed their judgments about the case. Given two essentially equivalent cases of reading disability, it was impossible to predict from one session on the case to the next session on the same case what diagnostic judgments any clinicians in this study would make about the case.

Statistical Agreement on Cue Categories

Cues from the categories of background and oral reading information were collected by clinicians in four of the eight sessions conducted on Case I, II, and III. This small clustering of relatively high frequency cues hints at a behavior pattern in the clinician's information gathering process. In at least 50 percent of the sessions per case, the clinicians apparently considered information contained within the cue categories of background and oral reading to be potentially relevant to their understanding of a case. It is of further interest to note the frequency with which the clinicians' diagnosis included judgments about background and oral reading.

The question is: "Were the diagnostic categories related to background and oral reading used as frequently as the cue categories providing information on background and oral reading?" The diagnostic category "Oral Reading Fluency - Weakness" was recorded in 50 percent of the sessions on Case II and III while judgments related to background were less frequent.

The mean intra-correlation coefficients on cue categories for Cases II and III were higher than the mean inter-correlation coefficients. The clinicians agreed a little more with themselves than with each other on the information to collect about Case II and Case III. Irrespective of instrumentation, the mean intra-correlation for Case II was .41 while the mean inter-correlation was .24; also, the mean intra-correlation for Case III was .35 while the mean inter-correlation was .16.

Statistical Agreement on Cue Categories Compared with Diagnostic Categories

Agreement on cue categories was relatively higher than agreement on diagnostic categories along all four statistical measures. The number of most frequently collected cues per case exceeded the number of most frequently recorded diagnostic categories per case. The average percentage of Proportional Agreement was approximately 30 percent on cue categories and 20 percent on

diagnostic categories. The inter-correlations on cue categories describing one clinician's agreement with another clinician were a little higher than the inter-correlations on diagnostic categories. Also, the intra-correlations on cue categories describing a clinician's agreement with him/her self were more distinctly higher than the intra-correlations on diagnostic categories.

The data described clinicians who generally tended to agree a little more on the type of information to examine when presented a case of reading disability than on diagnostic judgments about the case. This trend may be largely a result of the part-whole relationship of cues to diagnosis. In the problem-solving process, clinicians collect varying quantities of information in order to reduce uncertainty about a case and reach a diagnosis. It is possible that several cues regarding the reading behavior of the child represented in a case may have been chunked together by the clinician to formulate one diagnostic statement about the nature of the child's reading problem. It is also possible that this trend of higher agreement on cues than on diagnosis was caused in part by the cueing available. A Cue List was given to the clinicians assigned to PAPERCASES or SIM-CASES which listed all the items of information available

about a case. This list may have served as a stimulus that guided the clinicians' selection of information. While the Cue List was presented to the clinicians, the Diagnostic Checklist of diagnostic judgments was internally generated by the clinicians. The clinicians were given the Cue List to refer to before gathering information about the case, whereas they were given the Diagnostic Checklist after they diagnosed the case and asked to transfer their judgments to this checklist.

Methodology

As observational instruments the SIMCASES and PAPERCASES used in this study functioned as comprehensive clinical inventories of diagnostic performance as opposed to cognitive tests of diagnostic ability. Measures of agreement rather than measures of accuracy served as the criteria for describing diagnostic problem-solving behavior on these instruments. With the addition of a standardized scoring system these observational instruments could become assessment instruments capable of describing diagnostic performance in terms of how right or wrong, accurate or inaccurate a clinician's diagnostic judgments were in relation to the criteria of accuracy, the correct diagnosis of a case. If diagnosis performance was represented by an accuracy score or a thoroughness score or any other standardized scoring criteria,

then the universe of behavior would be more narrowly defined and traditional measurement principles could be more directly applied to the evaluation of the instruments.

Standardized scores would allow for a test-retest estimate of clinical reliability or the consistency over time of a clinician's diagnostic performance on the assessment instrument. A PAPERCASE test could be administered to a group of reading specialists then the same PAPERCASE test readministered at a later date, and an objective measure of the clinicians' reliability would be obtained from the correlation of the two sets of scores. Also, the equivalence of two forms of a PAPERCASE, an original and a replicate, could be more precisely estimated by administering the two forms of the test to the same group of individuals on the same day and correlating these results.

The concurrent validity of the PAPERCASE was suggested by the comparison of diagnostic performance on a SIMCASE and PAPERCASE with that on two SIMCASES. "In concurrent validity we are asking whether the test score can be substituted for some less efficient way of gathering criterion data" (Mehrens and Lehmann, 1973, p. 125). Although standardized scores were not used to measure diagnostic performance the fact that the

agreement statistics derived from the two instruments were very similar indicated that the PAPERCASE could be substituted for the less efficient SIMCASE as an observational instrument. Thus, the PAPERCASES appeared to possess concurrent validity. The development of a standardized scoring system would allow for more direct and precise estimates of validity or the degree to which the instruments measure what they purpose to measure.

Implications

The Problem-Solving Process of Diagnosis

The clinicians' low level of agreement with themselves and each other on the cues to select when presented with alternate forms of the same case of reading disability demonstrated a high level of inconsistency in the problem-solving process of diagnosis. The inconsistent cue collection behavior strongly suggests that the clinicians were randomly rather than systematically gathering information about the case. Although the Inquiry Theory states that clinical problem solving is determined probabilistically by the interaction of clinical memory, clinical strategy, and the case, in actual practice the clinicians in this study did not appear to have a strategy or a model in their minds of the diagnostic process. This observation has been suggested in previous observational studies

conducted by the Clinical Studies Project (OS77.1, OS78.2). The implication is that reading specialists need a practical model of reading diagnosis. Such a model would have considerable value if it were able to make explicit the abstract strategies and principles of diagnosis. If a model were able to elucidate these matters at a level of complexity which respected the constraints of human information processing, then reading specialists would be provided with a good clinical strategy or general framework with which to interact with a case of reading disability. The major question remaining is whether such a model can be formulated and taught to practitioners in reading.

Clinical Training in Diagnosis

The rather pessimistic results that clinicians most often disagree with themselves and each other on diagnostic judgments when presented with alternate forms of the same case of reading disability should not obscure the important implications of this study for training professionals in reading. According to Sherman (personal communication) who recently collaborated with the Clinical Studies Project in designing a graduate training course in Reading Diagnosis, "The good news is that we (reading specialists) can do much better than we do." In a

five-week training course recently taught by Sherman at Michigan State University the diagnostic ability of inservice teachers was improved by providing teachers with a Model of Reading and Learning to Read (Sherman, personal communication) and by providing the teachers with clinical experience, the opportunity to practice solving reading problems through the use of SIMCASES.

Within a graduate level course in Reading Diagnosis, PAPERCASES as well as SIMCASES could be utilized as alternatives to the face to face observation of a student diagnosing a reading problem. As simulation exercises, both instruments would give students the opportunity to practice solving a range of reading problems which closely approximate, in breadth and complexity, the problems encountered by reading specialists in the schools throughout this country.

With the addition of a feedback component, the PAPERCASES could efficiently and effectively be used as directed simulation exercises in diagnosing reading disabilities. Students could practice the skill of reading diagnosis on paper and evaluate their own growth, while the concepts and content of diagnosis becomes more explicit in this process of "learning by doing."

After demonstrating some level of initial mastery on the PAPERCASES students could then be assigned

to SIMCASES to further develop their skill in diagnosis. The SIMCASES could be used as undirected simulation exercises. Students could apply their skill to simulated cases which represented actual children with reading disabilities. Having successfully transferred their skill in diagnosis to a simulated child, the students would then have the prerequisite experience to begin diagnosing real children.

The students could move on in their training to a clinical practicum in Reading Diagnosis in which they would be assigned real cases to diagnose under the direct supervision of the clinical-faculty. The training emphsis within the practicum experience would be to increase accuracy and consistency in diagnosis.

Upon satisfactory completion of a clinical practicum, the students would be prepared for field work in reading diagnosis and assigned cases of reading disability in a school setting under the supervision of a competent reading specialist employed by the school system.

PAPERCASES may provide an invaluable practice component following initial instruction in how to diagnose reading disabilities and as independent and self-checking simulation exercises they may make a real difference in the learning outcome, the training of reliable reading clinicians who are consistently accurate in their diagnostic judgments.

The potential use of PAPERCASES in the training of reading specialists must be tempered by caution. Until the structure of abilities which comprise adequate problem solving in the area of reading diagnosis are clearly identified, student training and evaluation based on PAPERCASES and SIMCASES must remain qualified.

APPENDICES

APPENDIX A

INSTRUCTIONS FOR SIMCASE SESSIONS

APPENDIX A

INSTRUCTIONS

You have been asked to serve as a consultant for the Institute for Research on Teaching at Michigan State University in order that we may better understand how people like yourself diagnose and remediate children with reading problems.

Because this investigation takes several hours, the IRT will pay you as a consultant for the College of Education. A check will be mailed to you when your consulting work is completed. Please fill out this consultant pay form. (Give clinician pay form)

The Institute is required by law to protect your privacy by keeping confidential your name and social security number. Your name will not be part of Institute permanent records. Instead, a number will be used as identification.

Please read this Informed Consent Sheet. Your signature indicates that you agree to participate in the study and allow us to use the resulting information.

(Give time to read and sign Consent Sheet)

Please take a few minutes, now, to fill out this background information sheet. (Give clinician Background Information Sheet)

I would not like to give you an overview of the two sessions we will have together. In both sessions you will do two things. First, you will examine a simulated case of reading difficulty and will write a diagnostic and initial remediation plan. Second, you will focus on specific aspects of the written diagnosis and remediation.

Before we begin, we will practice using a simulated case of reading difficulty. Here is a file box in which is stored information about a child with reading difficulties. The information is available to you in five possible forms as shown on this cue inventory.

(Give clinician copy of inventory)

(1) Test scores; (2) Examiner's comments; (3) Test booklets; (4) Audio tapes, and (5) Test directions.

(Point to each category as it is being read)

The cue inventory for this practice case tells you specifically what information is available within each form. For example, the Durrell Oral Reading Test is available to you in all five forms:

(Show example of each form as it is being read)

As a test score; examiner's comments; test booklet; audio tape; and test directions.

You request information by referring to the inventory keyword. For example, if you want the results of the Durrell Oral Reading Test in the form of examiner's comments, you would get it by asking for the keyword DUR___. I would then hand you that piece of information. You may only request one item of information at a time.

This cue inventory is merely a listing of information that has already been collected for you. When you request specific forms of information, do so in the order you normally follow when working with a child.

You may now request two items of information using the inventory keywords. I will hand you the items you request.

(Give the clinician time to make the requests)

Having practiced using a case, you are now ready to request information about the case you will use this session. You will use this information to determine the most likely diagnosis and to suggest an initial remediation plan. Do not feel you must request an item of information just because it is present in the inventory; there is no right or wrong amount of information to request.

You may take notes if you wish and you may keep all forms of information requested until you have completed your work with the case.

After the presentation of some initial information about the case, you will have 30 minutes to collect information and reach a decision about the diagnosis. You will then have up to 20 minutes in which to write down your diagnosis and up to 20 minutes in which to write the initial remediation plan.

To review:

1. Collect information in the order you normally follow, using the keywords;
2. Take notes if you wish;
3. Keep the information requested if you wish;
4. Take up to 30 minutes to reach a decision about the diagnosis. I will let you know when 10 minutes remain.

Do you have any questions: (Pause)

You are now ready to begin examining a simulated case of reading difficulty. The case you are to consider today concerns a _____ year old boy named _____. Here is a sketch of the child, a statement of the reason for referral to the reading specialist, and a taped interview which I will now play for you. (Play tape)

I will start timing for 30 minutes. You may now request items of information.

BEGIN TAPE RECORDER
ENTER TIME/CUES ON CUE COLLECTION FORM
STOP AFTER 30 MINUTES

Would you now write your diagnosis? Please write complete sentences, not just key words or phrases. Assume that the report will be read by a clinician with training similar to yours. You will have up to 20 minutes to write your diagnosis. I will let you know when five minutes remain.

(Provide clinician with wide-lined paper)

Would you now write your initial remediation plan? Please write complete sentences, not just key words or phrases. Assume the plan will be used by a clinician with training similar to yours who will be responsible for working with the child. You will have up to 20 minutes to write your remediation plan. I will let you know when five minutes remain. Do you wish to keep the cues you've collected?

(Provide clinician with wide-lined paper)

BREAK

We are now going to begin the second part of our session during which you will focus on specific aspects of the written diagnosis and remediation.

The first step is for you to identify all of your key diagnostic elements. Here are a few statements from an abbreviated sample diagnosis. (Give sample diagnosis) This diagnosis is not meant to suggest a model; it merely provides practice in identifying types of diagnostic elements. Notice that the clinician circled what he or she considered to be key diagnostic elements and then numbered each circled element in turn. Notice, too, that the circle was left open at the end of the line when the element continued down to the next line. (Point out circled, numbered elements)

To complete the process, the clinician coded each diagnostic element as being either a strength, weakness, or observation, by placing an S for strength, W for weakness, or Obs for observation next to each circled, numbered element. (Point to codes) In contrast to a strength or weakness, an observation is simply a neutral statement that characterizes some aspect of the case.

I would now like you to practice identifying key diagnostic elements. First circle and number what you consider to be the remaining diagnostic elements in the sample diagnosis, and then code each one.

(Give time to complete "Practice" section)

I would now like you to do this task with your own written diagnosis. The sample diagnosis in no way implies that you should have a certain number of each type of element. It was designed merely to provide practice in identifying types of elements.

Use this red pen to circle and code the diagnostic elements.

(Circle, number, and code own diagnosis)

You will now match the circled elements to statements on this Diagnostic Checklist. Here is a sample checklist. (Give Dx Checklist) The checklist is divided into 12 major categories which appear as capitalized headings. (Point to categories and give clinician the separate table of contents sheet) This table of contents sheet will aid you in rapidly locating the categories.

The clinician who wrote the sample diagnosis decided that diagnostic element Number One fell under the main category FAMILY. (Point to appropriate heading) Under this category, the clinician further decided that element Number One was most similar to the sub-category 'Sibling Relationships." Since it had already been coded Obs for Observation, the clinician merely placed the number one under the appropriate column heading,

in this case Observation. (Point to the number (1) in the checklist) Please observe how the clinician transferred diagnostic elements 2 and 3.

(Give time to look at sample Dx and checklist)

Would you like to review the steps we've just gone through?

I would like you to practice matching your circled, numbered, and coded practice diagnostic elements to diagnostic elements in the sample checklist. (Give time to match)

Here is another copy of the Diagnostic Checklist and your written diagnosis. Please match your written diagnostic elements to the checklist according to the steps we just reviewed. If you have a diagnostic element that does not correspond to any of the main categories, list it under the main category OTHER. (Point out) If you have a diagnostic element that corresponds to a main category but does not correspond to any subcategory under it, list it under the appropriate "Other" category. (Point out) Please make as limited use as possible of these "Other" categories.

You will not be timed. Please work as rapidly as is comfortable. You may begin.

(When completed, make sure all the written Dx numbers are accounted for in the DX Checklist)

Please turn your attention now to your remediation plan. You will be asked to identify all of your key remedial elements. Here are a few statements from an abbreviated sample remediation. This remediation is not meant to suggest a model. It merely provides practice in identifying types of remedial elements. Notice that the clinician circled what she or he considered to be the key remedial elements and then numbered each circled element in turn. Notice that the circle was left open at the end of the line whenever the element continued down to the next line. (Point out circled, numbered elements) To complete the process, the clinician coded each remedial element as being either a strength, weakness, observation, or treatment, by placing an S for strength, W for weakness, Obs for observation or T for treatment next to each circled, numbered element. (Point to coded elements)

I would now like you to practice identifying key remedial elements. First circle and number, and then code what you consider to be the remaining remedial elements in the sample remediation.

(Give time to complete "Practice" section of remediation)

I would now like you to do this task with your own written remediation. The sample remediation in no way implies that you should have a certain number of each type of element. It was designed solely to provide practice in identifying types of elements.

Use this red pen to circle, number and code the remedial elements.

(Ciecle, number, and code own remediation)

You will not match the circled elements to statements on this remedial checklist. Here is a sample checklist. (Give Rx checklist)

This sample checklist is divided into 12 major categories which appear as capitalized headings. (Point to categories and give clinician the separate table of contents sheet) This table of contents sheet will aid you in rapidly locating the categories.

The clinician who wrote the sample remediation decided that remedial element Number One fell under the main category VISUAL. (Point to appropriate heading) Under this main category, the clinician further decided that element Number One was most similar to the subcategory "Discriminate Visually Similar Words." Since it had already been coded T for Treatment, the clinician merely placed the number (1) under the appropriate

column heading, in this case T. (Point to the number (1) in the checklist) In this particular instance, there was more in the written element than was present on the checklist. The clinician wrote in the additional comment, "Highlight the middle." (Point to written-in comment) Please notice how the clinician transferred remedial ements 2, 3, and 4.

(Give time to look at sample Rx and checklist)

Would you like to review the steps we've just gone through?

I would like you to practice matching your circled, numbered, and coded practice remedial elements to remedial elements in the sample checklist. (Give time to match)

Here is another copy of the Remedial checklist and your written remediation. I would like for you to match your written remedial elements to the checklist according to the steps we just reviewed. Remember: You can add additional remedial comments in the space provided. (Point) If you have a remedial element that does not correspond to any of the main categories, list it under the main category labelled OTHER. (Point out) If you have a remedial element that corresponds to a main category but does not correspond to any sub-category

under it, list is under the appropriate "Other" category. (Point out) Please make as limited use as possible of these "Other" categories. You will not be timed. Please work as rapidly as is comfortable.

(When clinician has finished remedial checklist, take both the diagnostic and remedial checklists and file. Give clinician his/her written diagnosis. Exp. keeps written Rx)

APPENDIX B

CUE LISTS FOR SIMCASE SESSIONS

APPENDIX B

Case I

CASE INFORMATION INVENTORY

FORMS OF INFORMATION

TYPES OF INFORMATION	TEST SCORES		TEST BOOKLET			TEST DIRECTIONS		
		EXAMINER'S COMMENTS			AUDIO RECORDING OF SESSION			
NON-STANDARDIZED ASSESSMENT								
BASIC SIGHT VOCABULARY (DOLCH WORD LIST)	DOL1	DOL2	DOL3			DOL4		
SENTENCE COMPLETION		SENT2	SENT3	SENT4		SENT5		
GATES MCKILLOP READING DIAGNOSTIC TESTS								
RECOG. AND BLENDING COMMON WORD PARTS	GMK1	GMK2	GMK3	GMK4		GMK5		
AUDITORY BLENDING	GMK7	GMK8	GMK9	GMK10		GMK11		
GIVING LETTER SOUNDS	GMK13	GMK14	GMK15	GMK16		GMK17		
AUDITORY DISCRIMINATION (WEPMAN AUDITORY DISCRIM. TEST)	WEP1	WEP2	WEP3			WEP5		
PHYSICAL INFORMATION								
VISION TEST		VIS2	VIS3					
AUDIOMETRIC RECORD		AUD2	AUD3					
BACKGROUND INFORMATION								
SCHOOL RECORD			BKG15					
TEACHER FORM			BKG9					
SCHOOL INFORMATION			BKG19					
PARENT FORM			BKG3					
STANDARDIZED ASSESSMENT								
DURRELL LISTENING-READING SERIES INTERMEDIATE LEVEL								
VOCABULARY	DLR1	DLR2	DLR3			DLR5		
PARAGRAPHS	DLR7	DLR8	DLR9					

APPENDIX B

Case I

CASE INFORMATION INVENTORY

FORMS OF INFORMATION

TYPES OF INFORMATION	TEST SCORES	EXAMINER'S COMMENTS	TEST BOOKLET	AUDIO RECORDING OF SESSION	TEST DIRECTIONS	
NON-STANDARDIZED ASSESSMENT						
BASIC SIGHT VOCABULARY (DOLCH WORD LIST)	DOL1	DOL2	DOL3		DOL5	
SENTENCE COMPLETION		SENT2	SENT3	SENT4	SENT5	
GATES MCKILLOP READING DIAGNOSTIC TESTS						
RECOG. AND BLENDING COMMON WORD PARTS	GMK1	GMK2	GMK3	GMK4	GMK5	
AUDITORY BLENDING	GMK7	GMK8	GMK9	GMK10	GMK11	
GIVING LETTER SOUNDS	GMK13	GMK14	GMK15	GMK16	GMK17	
AUDITORY DISCRIMINATION (WEPMAN AUDITORY DISCRIM. TEST)	WEP1	WEP2	WEP3		WEP5	
PHYSICAL INFORMATION						
VISION TEST		VIS2	VIS3			
AUDIOMETRIC RECORD		AUD2	AUD3			
BACKGROUND INFORMATION						
SCHOOL RECORD			BKG18			
TEACHER FORM			BKG9			
SCHOOL INFORMATION			BKG19			
PARENT FORM			BKG3			
STANDARDIZED ASSESSMENT						
DURRELL LISTENING-READING SERIES INTERMEDIATE LEVEL						
VOCABULARY	DLR1	DLR2	DLR3		DLR5	
PARAGRAPHS	DLR7	DLR8	DLR9			

TYPES OF INFORMATION	FORMS OF INFORMATION: TEST SCORES		TEST BOOKLET		TEST DIRECTIONS	
		EXAMINER'S COMMENTS		AUDIO RECORDING OF SESSION		
DIAGNOSTIC ANALYSIS OF READING DIFFICULTY (DURRELL)						
ORAL READING	DUR1	DUR2	DUR3	DUR4	DUR5	DUR6
SILENT READING	DUR7	DUR8	DUR9	DUR10	DUR11	DUR12
LISTENING COMPREHENSION	DUR13	DUR14	DUR15		DUR17	DUR18
WORD RECOGNITION AND WORD ANALYSIS	DUR19	DUR20	DUR21	DUR22	DUR23	DUR24
HEARING SOUNDS IN WORDS-PRIMARY	DUR25	DUR26	DUR27		DUR29	DUR30
VISUAL MEMORY OF WORDS-PRIMARY	DUR31	DUR32	DUR33		DUR35	DUR36
INTERMEDIATE SPELL. LIST 1	DUR37	DUR38	DUR39		DUR41	DUR42
PHONIC SPELLING OF WORDS	DUR43	DUR44	DUR45		DUR47	DUR48
ACHIEVEMENT TEST (IOWA TEST OF BASIC SKILLS)						
VOCABULARY SUBTEST	ITBS1	ITBS2	ITBS3		ITBS5	ITBS6
READING SUBTEST	ITBS7	ITBS8	ITBS9		ITBS6	ITBS6
GRADED WORD LIST (GLOSSON ORAL READING TEST)	SORT1	SORT2	SORT3		SORT5	SORT6
READING ACHIEVEMENT (GATES MACGINITIE) SPEED/ACCURACY TEST	GMG1		GMG3		GMG5	
COGNITIVE ABILITY (WECHSLER INTELLIGENCE SCALE FOR CHILDREN WISC-R)						
VERBAL SCALE	WISR1				WISR5	
PERFORMANCE SCALE	WISR7				WISR11	
FULL SCALE	WISR13	WISR14	WISR15			

Case II

CASE INFORMATION INVENTORY

TYPES OF INFORMATION	FORMS OF INFORMATION: TEST SCORES	TEST BOOKLET: EXAMINEE'S COMMENTS	TEST BOOKLET: AUDIO RECORDING OF SESSION	TEST DIRECTIONS: TUTORIAL
STANDARDIZED ASSESSMENT				
GRADED WORD LIST (SLOSSON ORAL READING TEST)	SORT1	SORT2	SORT3	SORT4	SORT5
COGNITIVE ABILITY(WISC)					
FULL SCALE	WISC1	WISC2			
VERBAL SCALE	WISC7	WISC8			WISC11
PERFORMANCE SCALE	WISC13	WISC14			WISC17
DIAGNOSTIC TEST OF READING DIFFICULTY-(GATES-MCKILLOP)					
RECOGNIZING AND BLENDING COMMON WORD PARTS	GMK1	GMK2	GMK3	GMK4	GMK5
ACHIEVEMENT TEST-INDIVIDUAL (PEABODY)					
READING RECOGNITION	PEA1	PEA2	PEA3		PEA5
READING COMPREHENSION	PEA7	PEA8	PEA9		PEA5
SPELLING	PEA13	PEA14	PEA15		PEA5
GENERAL INFORMATION	PEA19		PEA21		PEA6
ACHIEVEMENT TEST-GROUP(GATES-MACGINITIE)					
VOCABULARY	GMG1	GMG2	GMG3		
COMPREHENSION	GMG7	GMG8	GMG9		
SPEED AND ACCURACY	GMG13		GMG15		
DIAGNOSTIC TEST OF READING DIFFICULTY (DURRELL)					
ORAL READING	DUR1	DUR2	DUR3	DUR4	DUR5
SILENT READING	DUR7	DUR8	DUR9		DUR11
LISTENING COMPREHENSION	DUR13	DUR14	DUR15		DUR17
WORD RECOGNITION AND WORD ANALYSIS	DUR19	DUR20	DUR21	DUR22	DUR23
VISUAL MEMORY OF WORDS-PRIMARY	DUR25	DUR26	DUR27		DUR29
HEARING SOUNDS IN WORDS	DUR31	DUR32	DUR33		DUR35
SOUNDS OF LETTERS	DUR37	DUR38	DUR39		

TYPES OF INFORMATION	FORMS OF INFORMATION: TEST SCORES	TEST BOOKLET: EXAMINER'S COMMENTS	TEST BOOKLET	TEST BOOKLET: AUDIO RECORDING OF SESSION	TEST DIRECTIONS	TEST DIRECTIONS: TUTORIAL
BACKGROUND INFORMATION						
PHYSICAL/HEALTH		BKG8				
HOME/FAMILY		BKG14				
CLASSROOM INFORMATION		BKG20				
BIOGRAPHICAL DATA		BKG2				
NON STANDARD ASSESSMENT						
INDIVIDUAL READING SUBSKILL ANALYSIS (SOUND-SYMBOL ASSOC.)			IPA3		TPA5	
INFORMAL ORAL READING		INF2		INF4	INF5	
BASIC SIGHT VOCABULARY (COLOR)	DOL1	DOL2	DOL3	DOL4	DOL5	

Case III

CASE INFORMATION INVENTORY

TYPES OF INFORMATION	FORMS OF INFORMATION: TEST SCORES	EXAMINER'S COMMENTS	TEST BOOKLET	AUDIO RECORDING OF SESSION	TEST DIRECTIONS	TUTORIAL
PHYSICAL INFORMATION						
AUDITORY ACUITY		AUD2	AUD3			
VISUAL ACUITY		VIS2	VIS3			
STANDARDIZED ASSESSMENT						
COGNITIVE ABILITY (SLOSSON INTELLIGENCE TEST)	SIT1	SIT2	SIT3		SIT5	
ACHIEVEMENT TEST GROUP-ARITHMETIC (ARITHMETIC COMPUTATION)	ARI1	ARI2	ARI3		ARI5	
COGNITIVE ABILITY (WISC-R)						
FULL SCALE	WISR1	WISR2	WISR3			
VERBAL SCALE	WISR7	WISR8	WISR3		WISR11	
PERFORMANCE SCALE	WISR13	WISR14	WISR3		WISR17	
ACHIEVEMENT TEST GROUP-READING (GATES-MACGINITIE)						
VOCABULARY	GMG1	GMG2	GMG3		GMG5	
COMPREHENSION	GMG7	GMG8	GMG9		GMG11	
SPEED AND ACCURACY	GMG13	GMG14	GMG15		GMG17	
DIAGNOSTIC TEST OF READING DIFFICULTY (GATES-MCKILLOP) RECOGNIZING AND BLENDING COMMON WORD PARTS	GMK1	GMK2	GMK3		GMK5	
NONSENSE WORDS	GMK7	GMK8	GMK9		GMK11	
SYLLABICATION	GMK13	GMK14	GMK15		GMK17	
AUDITORY DISCRIMINATION (SPACHE DIAGNOSTIC READING SCALES - AUDITORY DISCRIMINATION)	SPA1	SPA2	SPA3	SPA4	SPA5	
GRADED WORD LIST (SLOSSON ORAL READING						

TYPES OF INFORMATION	FORMS OF INFORMATION: TEST SCORES	EXAMINER'S COMMENTS	TEST BOOKLET	AUDIO RECORDING OF SESSION	TEST DIRECTIONS	TUTORIAL
DIAGNOSTIC TEST OF READING DIFFICULTY (DURRELL)						
ORAL READING	CUR1	CUR2	CUR3	CUR4	CUR5	
SILENT READING	CUR7	CUR8	CUR9		CUR11	
LISTENING						
COMPREHENSION	CUR13	CUR14	CUR15		CUR17	
WORD RECOGNITION AND						
WORD ANALYSIS	CUR19	CUR20	CUR21		CUR23	
VISUAL MEMORY OF						
WORDS-INTERMEDIATE	CUR25	CUR26	CUR27		CUR29	
SPELLING	CUR31	CUR32	CUR33		CUR35	
PROFILE CHART	CUR37					
NON-STANDARDIZED ASSESSMENT						
STRUCTURAL ANALYSIS INVENTORY						
WRITTEN	SAI1	SAI2	SAI3		SAI5	
ORAL	SAI7	SAI8	SAI9			
HANDWRITING		HAN2	HAN3		HAN5	
ATTITUDE (UNFINISHED SENTENCES)			UNF3	UNF4		
SYLLABICATION BLENDING	SYL1	SYL2	SYL3	SYL4	SYL3	
COMPREHENSION (INDIVIDUAL READING SKILLS PROGRAMS, LEVEL B)	IRS1	IRS2	IRS3		IRS5	
INDIVIDUAL READING ANALYSIS CONSONANT BLENDS	IRA1	IRA2	IRA3		IRA3	
BACKGROUND INFORMATION						
SCHOOL RECORD		BKG14				
PARENT FORM		BKG2				
STANDARDIZED ACHIEVEMENT TEST SCORES	BKG25		BKG27			
TEACHER FORM		BKG8				
PUPIL PROGRESS REPORTS		BKG20				

APPENDIX C

DIAGNOSTIC CHECKLIST FOR ALL SESSIONS

APPENDIX C

DIAGNOSTIC CHECKLIST

Clinical Studies Project
Institute for Research on Teaching
Michigan State University

Clinician: _____

Session: _____

Case: _____

Date: _____

Time: _____

Location: _____

S = Strength; W = Weakness; O = Observation

	S	W	O		S	W	O
PHYSICAL				FAMILY			
Physical: General				Family: General			
Physical development				Home environment			
General health				Parent-school cooperation			
Vision: General				Parent educational background			
Vision: Acuity				Parent attitude toward school			
Hearing: General				Parent-child relationship			
Hearing: Acuity				Sibling relationship			
Speech articulation				Other			
Motor coordination							
Other							
				PERCEPTION			
				Perception: General			
				Auditory memory			
				Auditory discrimination			
				Auditory sequencing			
AFFECTIVE				Auditory blending			
Affective: General				Visual memory			
Attitude toward school				Visual discrimination			
Attitude toward reading: Instructional				Visual sequencing			
Attitude toward reading: Independent				Visual motor integration			
Relationships with peers				Other			
Motivation for reading							
Behavior in the classroom							
Emotional adjustment				LANGUAGE			
Self-concept				Language: General			
Variety of interests				Listening vocabulary			
Ability to deal with new situations				Speaking vocabulary			
Other				Verbal fluency			
				Other			

S = Strength; W = Weakness; O = Observation 2

	S	W	O		S	W	O
SCHOOL				ORAL READING			
School: General				Oral reading: General			
Grade level placement				Oral reading: Rate			
Ability to work in a tutoring situation				Oral reading: Fluency			
Ability to work in a small group				Oral reading: Punctuation			
Ability to apply reading skills				Oral reading: Self-correction			
Attending behavior				Oral reading: Phrasing			
Intellectual potential: General				Oral reading: Intonation			
Intellectual potential: Verbal				Substitutions contextually acceptable			
Intellectual potential: Nonverbal				Omissions contextually acceptable			
Ability to do grade level work				Insertions contextually Acceptable			
Ability to read at grade level				Use of context to determine word pronunciation			
Appropriateness of instructional materials				Independent reading level: Fluency			
Other				Instructional reading level: Fluency			
				Other			
STRUCTURAL ANALYSIS				SILENT READING			
Word Analysis: General				Silent Reading: General			
Structural Analysis: General				Silent Reading: Rate			
Use of root words				Silent Reading: Comprehension			
Use of suffixes				Other			
Use of prefixes							
Use of syllables							
Blending word parts into whole words							
Other							

S = Strength; W = Weakness; O = Observation

3

	S	W	O		S	W	O
PHONETIC ANALYSIS				WORD RECOGNITION			
Word analysis: General				Word recognition: General			
Phonetic analysis: General				Word recognition: Basic sight words			
Use of letter-sound association				Use of whole word approach			
Use of initial consonant sounds				Use of initial letters in word identification			
Use of final consonant sounds				Use of final letters in word identification			
Use of blends: General				Use of medial letters in word identification			
Use of consonant diagraphs				Use of letter order (reversals)			
Use of word families				Other			
Use of syllables							
Ability to blend sounds into whole word							
Other							
				COMPREHENSION			
				Comprehension: General			
				Comprehension: Oral			
				Comprehension: Silent			
				Comprehension: Listening			
OTHER				Comprehension: Vocabulary			
				Use of context to get meaning			
				Use of word order to get meaning			
				Recall of information			
				Comprehension: Use of Specific strategies			
				Other			

APPENDIX D

INSTRUCTIONS FOR PAPERCASE SESSIONS

APPENDIX D

READING DIAGNOSIS PROBLEM

CASE: Andy

CLINICAL STUDIES PROJECT

INSTITUTE FOR RESEARCH ON TEACHING

MICHIGAN STATE UNIVERSITY

SUMMARY

The diagnostic interaction you are about to engage in is divided into the following 6 subtasks:

1. You will be given a maximum of 30 minutes to select and record the information you need from the CASE NOTEBOOK to diagnose the reading problem.
2. You will be given a maximum of 20 minutes to write your diagnosis.
3. You will be given a maximum of 20 minutes to write your remediation plan.
4. The group will take a short break.
5. You will transfer your diagnostic statements to a standardized Diagnostic Checklist.
6. You will transfer your remediation statements to a standardized Remediation Checklist.

Please turn to the INITIAL CONTACT information which will introduce you to your CASE.

INSTRUCTIONS
READING DIAGNOSIS PROBLEM

This notebook contains a case which represents a child with a reading problem. The information which has been collected about this case is listed in the CASE INFORMATION INVENTORY. Please remove and unfold the CASE INFORMATION INVENTORY (cream colored sheet) from the front jacket pocket of this notebook. You will continually refer to this inventory as you need to select forms of information about the case.

Notice the first column of the CASE INFORMATION INVENTORY is labeled SOURCE OF INFORMATION. Under this heading various sources of potentially important information are listed. Column 2 is labeled TEST BOOKLET AND BEHAVIOR, Column 3 is TEST SCORES, Column 4 is TEST DIRECTIONS and Column 5 is labeled OTHER DATA RECORDS. Under these columns, the different forms of information are specified as well as their location. Information in this notebook is referenced through the use of a letter and symbol system.

As you select a form of information, record the letter or symbol which represents the information on the CASE INFORMATION LIST. Please remove the sample list (green colored sheet) from the front jacket pocket of this notebook. For example, if you want to examine the results of a test of SIGHT VOCABULARY and select the Dolch Word List as your source of information and the TEST BOOKLET AND BEHAVIOR as the form of information, then you would:

1. Print the upper case or capitol letter "A" on the first line of the CASE INFORMATION LIST.
2. Locate the capitol "A" tab in the notebook.

3. Turn to the page tabbed capitol "A"
4. Examine the TEST BEHAVIOR (note: many forms of information extend over more than one page in this notebook)
5. Then return to the CASE INFORMATION INVENTORY and select your next piece of data.

If you select TEST SCORES on the Dolch Word List as your next form of data notice that you record the lower case or small letter "a" on the next line of the CASE INFORMATION LIST. Notice that if you select TEST DIRECTIONS as a form of data, you record a plus sign "+" on the nest line of the CASE INFORMATION LIST. Whatever form of information you select you are required to <u>first</u> record the letter or symbol for the form on the CASE INFORMATION LIST before referring to the information.

You may select as many or as few sources of information as seem appropriate to diagnose the reading problem. The fact that the information is available does not imply that it is necessary for your understanding of the case. Examine only the information you regard as pertinent. Follow your usual information gathering procedures. Feel free to take notes if you wish on the writing tablet provided in the back of this notebook.

You will be given a maximum of 30 minutes to examine information on this case and reach a decision about the diagnosis. Then, you will be given a maximum of 20 minutes to write your diagnosis on the pink paper found in the front jacket of this notebook. Then you will be given another 20 minutes maximum to write an initial remediation plan on the blue form found in the front jacket of this notebook. After a short break, you will transfer your diagnostic statements to a Diagnostic Checklist and your remediation statements to a Remediation Checklist.

APPENDIX E

CUE LISTS FOR PAPERCASE SESSIONS

APPENDIX E

READING DIAGNOSTIC PROBLEM
INFORMATION INVENTORY
CASE: BRIAN

SOURCE OF INFORMATION	FORM OF INFORMATION			
	TEST BOOKLET AND BEHAVIOR	TEST SCORES	TEST DIRECTIONS	OTHER DATA RECORDS
TESTS OF SIGHT VOCABULARY				
Dolch: Word List	A	a	+	
Durrell: Flash/Analysis Subtest	B	b	+	
Slossan: Oral Reading Test	C	c	+	
STANDARDIZED READING ACHIEVEMENT TESTS				
Gates McGinitie: Speed & Accuracy Subtest	D	d	+	
Iowa Test of Basic Skills: Comprehension Subtest	E	e	+	
Iowa Test of Basic Skills: Vocabulary Subtest	F	f	+	
TESTS OF PHONETIC ABILITY				
Durrell: Hearing Sounds in Words Subtest	G	g	+	
Durrell: Phonetic Spelling of Words Subtest	H	h	+	
Durrell: Spelling Subtest -List 1	I	i	+	
Gates McKillop: Auditory Blending Subtest	J	j	+	
Gates McKillop: Giving Letter Sounds Subtest	K	k	+	
Gates McKillop: Recognizing and Blending Common Word Parts (Modified)	L	l	+	
TESTS OF ORAL READING & COMPREHENSION				
Durrell: Oral Reading Subtest	M	m	+	
Informal Basal Oral Reading	N	n	+	
TEST OF SILENT READING & COMPREHENSION				
Durrell: Silent Reading Subtest	O	o	+	
TESTS OF LISTENING COMPREHENSION				
Durrell: Listening Comprehension Subtest	P	p	+	
Durrell: Listening Test: Vocabulary Subtest	Q	q	+	
Durrell: Listening Test: Paragraphs Subtest	R	r	+	
TESTS OF LANGUAGE OR COGNITIVE POTENTIAL				
WISC-R: Full Scale		s		
WISC-R: Performance Scale	S	t	+	
WISC-R: Verbal Scale	S	u	+	
TEST OF VISUAL PERCEPTION				
Durrell: Visual Memory of Words-Primary	V	v	+	
TEST OF AUDITORY DISCRIMINATION				
Wepman: Auditory Discrimination Test	W	w	+	
PHYSICAL HEALTH RECORDS				
Audiometric Record	X	x		
Vision Test Record	Y	y		
BIOGRAPHICAL RECORDS				
Home/Family Records			Z	
School/Classroom Records			AA	
Attitude Inventory			BB	

READING DIAGNOSTIC PROBLEM
INFORMATION INVENTORY

CASE: ANDY

SOURCE OF INFORMATION	FORM OF INFORMATION			
	TEST BOOKLET AND BEHAVIOR	TEST SCORES	TEST DIRECTIONS	OTHER DATA RECORDS
TESTS OF SIGHT VOCABULARY				
Dolch: Basic Sight Word List	A	a	+	
Durrell: Flash Analysis	B	b	+	
Slossan: Oral Reading Test	C	c	+	
STANDARDIZED READING ACHIEVEMENT				
Gates McGinitie: Comprehension Subtest	D	d	+	
Gates McGinitie: Speed and Accuracy Subtest	E	e	+	
Gates McGinitie: Vocabulary Subtest	F	f	+	
Peabody: Comprehension Subtest	G	g	+	
Peabody: General Information Subtest	H	h	+	
Peabody: Reading Recognition Subtest	I	i	+	
Peabody: Spelling Subtest	J	j	+	
TESTS OF PHONETIC ABILITY				
Durrell: Hearing Sounds in Words Subtest	L	l	+	
Durrell: Sounds of Letters Subtest	M	m	+	
Gates McKillop: Recognizing and Blending Common Word Parts Subtest (Modified)	K	k	+	
TESTS OF ORAL READING & COMPREHENSION				
Durrell: Oral Reading Subtest	N	n	+	
Informal Basal Oral Reading	O	o	+	
TEST OF SILENT READING & COMPREHENSION				
Durrell: Silent Reading Subtest	P	p	+	
TESTS OF LANGUAGE OR COGNITIVE POTENTIAL				
Durrell: Listening Comprehension Subtest	Q	q	+	
WISC-R: Full Scale		r		
WISC-R: Performance Scale		s	+	
WISC-R: Verbal Scale		t	+	
TEST OF VISUAL PERCEPTION				
Durrell: Visual Memory Subtest; Primary	U	u	+	
TEST OF AUDITORY DISCRIMINATION				
Wepman: Auditory Discrimination Test	V	v	+	
PHYSICAL HEALTH RECORDS				
Audiometric Record	W	w		
Vision Test Record	X	x		
BIOGRAPHICAL RECORDS				
Classroom Information				Y
Home/Family Record				Z
Physical/Health Record				aa

READING DIAGNOSTIC PROBLEM
INFORMATION INVENTORY

CASE: ALLEN

SOURCE OF INFORMATION	FORM OF INFORMATION			
	TEST BOOKLET AND BEHAVIOR	TEST SCORES	TEST DIRECTIONS	OTHER DATA RECORDS
TESTS OF SIGHT VOCABULARY				
Durrell: Flash Analysis	A	a	+	
Slossan: Oral Reading Test	B	b	+	
STANDARIDZED ACHIVEMENT TESTS				
California Test of Basic Skills:				
Reading Subtest	C	c	+	
Language Subtest	D	d	+	
Study Skills Subtest	E	e	+	
Michigan Educational Assessment Test:				
Math Subtest	F	f	+	
Reading Subtest	G	g	+	
STANDARDIZED READING ACHIEVEMENT TESTS				
Gates McGinitie: Comprehension Subtest	H	h	+	
Gates McGinitie: Vocabulary Subtest	I	i	+	
Gates McGinitie: Speed and Accuracy Subtest	J	j	+	
TESTS OF PHONETIC ABILITY				
Durrell: Spelling Subtest	K	k	+	
Gates McKillop: Nonsense Words Subtest	L	l	+	
Gates McKillop: Recognizing and Blending Common Word Parts Subtest (Modified)	M	m	+	
Gates McKillop: Syllabication Subtest	N	n	+	
San Diego: Syllabication Blending Subtest (Adapted)	O	o	+	
TEST OF STRUCTURAL ANALYSIS ABILITY				
Structural Analysis Inventory Parts I & II	P	p	+	
TESTS OF ORAL READING AND COMPREHENSION				
Durrell: Oral Reading Subtest	Q	q	+	
Informal Basal Oral Reading	R	r	+	
TESTS OF SILENT READING AND COMPREHENSION				
Durrell: Silent Reading Subtest	S	s	+	
Individual Reading Skills Programs, Level B	T	t	+	
TESTS OF LANGUAGE OR COGNITIVE POTENTIAL				
Durrell: Listening Comprehension Subtest	U	u	+	
WISC-R: Full Scale	V	v	+	
WISC-R: Performance Scale	W	w	+	
WISC-R: Verbal Scale	X	x	+	
TEST OF VISUAL PERCEPTION				
Durrell: Visual Memory of Words, Intermediate Level	Y	y	+	
TEST OF AUDITORY DISCRIMINATION				
Wepman: Auditory Discrimination Test	Z	z	+	
PHYSICAL RECORDS				
Auditory Acuity Record	AA	aa	+	
Visual Acuity Record	BB	bb	+	
BIOGRAPHICAL RECORDS				
Attitude Inventory				cc
Parent/Home Record				dd
Pupil Progress Record				ee
School Record				ff
Teacher/Classroom Record				gg

REFERENCES

REFERENCES

Aaron, J. E. What teachers and prospective teachers know about phonic generalizations. Journal of Educational Research, 1960, 53, 323-330.

Artley, A. S. and Hardin, V. Inventory of teacher knowledge of reading (revised edition). Columbia, Missouri: Lucas Brothers Publishers, 1975.

Bond, G. L. and Tinker, M. A. Reading difficulties, their diagnosis and correction (third edition). New York: Appleton-Century-Crofts, Meredith Corporation, 1973.

Browman, B. L. Factors associated with teacher knowledge of reading skills. Unpublished Ph.D. dissertation, University of Torrence, 1962.

Bruecker, L. J. Introduction, educational diagnosis. In E. Dechant, Diagnosis and remediation of reading disabilities. West Nyack, New York: Parker Publishing Company, Inc., 1968, 3.

Bruner, J. S. Toward a theory of instruction. Cambridge, Mass.: Belknap Press, 1966.

Burnett, R. W. The diagnostic problem solving proficiency of elementary school teachers in teaching reading. Unpublished Ph.D. dissertation, University of Indiana, 1961.

Carroll, J. B. Words, meanings and concepts. Harvard Educational Review, 1964, 34, 2, 178-202.

Carter, H. L. J. and McGinnis, D. J. Diagnosis and treatment of the disabled reader. London: MacMillian Company, 1970.

Cruickshank, D. R. and Kennedy, J. J. The institute for research on teaching: an overview. Journal of Teacher Education, 1977, 28, 4, 44-49.

Dechant, E. Diagnosis and remediation of reading disability. West Nyack, New York: Parker Publishing Company, Inc., 1968.

Della-Piana, G. M. Reading diagnosis and prescription: an introduction. New York: Holt, Rinehart, and Winston, 1968.

Dewey, J. Logic: the theory of inquiry. New York: Holt, 1938.

Durkin, D. Phonics test for teachers. New York: Teachers College Press, 1964.

Ekwall, E. E. Diagnosis and remediation of the disabled reader. Boston: Allyn and Bacon, Inc., 1976.

Elstein, A. S., Shulman, L. S., and Sprafka, S. Medical problem solving: an analysis of clinical reasoning. Cambridge, Mass.: Harvard University Press, 1978.

Elstein, A. S., Shulman, L. S., Vinsonhaler, J. F. Wagner, C. C., and Bader, L. Characteristics of the clinical problem solving model and its relevance to educational research. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Elstein, A. S., Sprafka, S. A., and Shulman, L. S. Analyzing Medical Inquiry Processes. Paper presented at 1973 Annual Meeting of American Educational Research Association, New Orleans, LA., February 1973.

Emans, R. Teacher evaluations of reading skills and individualized reading. The Elementary English Review, 1965, 42, 258-260.

Gagné, R. M. Instruction based on research in learning. Engineering Education, 1971, 61, 519-523.

Gil, D., Hoffmeyer, L., Van Roekel, J., and Weinshank, A. Clinical problem solving in reading: theory and research (Research Series Number 45). The Institute for Research on Teaching, Michigan State University, 1979.

Gil, D., Vinsonhaler, J. F., and Sherman, G. Defining reading diagnosis: why, what, and how? (Research Series Number 46). The Institute for Research on Teaching, Michigan State University, 1979.

Gray, W. S. Remedial cases in reading, their diagnosis and treatment. Supplementary education monographs, No. 22, Department of Education. University of Chicago, 1922.

Guthrie, J. T. Graduate programs and faculty in reading. Newark, Delaware: International Reading Association, 1976.

International Reading Association Professional Standards and Ethics Committee. Roles, responsibilities and qualifications of reading specialists (a brochure). Newark, Delaware: International Reading Association, 1979.

Karlsen, B. Reading: assessment and diagnosis of abilities. In P. Lamb and R. Arnold (eds.), Reading: foundations and instructional strategies. Belmont, California: Wadsworth Publishing Company, Inc., 1976, 147-191.

Kennedy, E. C. Classroom approaches to remedial reading. Itasca, Illinois: F. E. Peacock Publishers, Inc., 1971.

Kingston, A. J. The inventory of teacher knowledge of reading--a validation. Reading Teacher, 1975, 29, 133-136.

McGuigan, F. J. Experimental psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

Mehrens, W. A. and Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart, and Winston, Inc., 1973.

Monroe, M. and Backus, B. Remedial reading. Cambridge, Mass.: Riverside Press, 1937.

Porter, A. Personal communication. Michigan State University, January 2 and 26, 1979.

Professional Standards and Ethics Committee of the International Reading Association. Guidelines for the professional preparation of reading teachers. Reading Teacher, 1978, 32, 1, 48-55.

Ramsey, W. Z. Will tomorrow's teachers know and teach phonics. Reading Teacher, 1962, 15, 242-245.

Rutherford, W. L. From diagnosis to treatment of reading disabilities. Academic Therapy Quarterly, 1972, 8, 1, 51-55.

Schubert, D. G. Teachers and word analysis skills. Journal of Reading, 1959, 12, 62-64.

Sherman, G. Personal communication. Michigan State University, Fall, 1979.

Sherman, G. Personal communication. Michigan State University, Spring, 1979.

Smith, H. P. and Dechant, E. V. Psychology in teaching reading. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962.

Spache, G. D. Diagnosis and remediation in 1980. In the Proceedings of the International Reading Association Conference. Newark, Delaware: International Reading Association, 1968, 13 Part 3, 135-151.

Spache, G. D. Diagnosing and correcting reading disabilities. Boston: Allyn and Bacon, Inc., 1976.

Spache, G. and Baggett, R. What do teachers know about phonics and syllabication. Reading Teacher, 1965, 19, 96-99.

Stein, E., Hauck, L. C., and Su, P. Y. (eds.). The random house dictionary of the english language. New York: Random House, 1973.

Strang, R. Diagnostic teaching of reading. New York: McGraw-Hill, 1964.

Strang, R. Reading diagnosis and remediation. Newark, Delaware: International Reading Association Monograph Series, 1968.

Strang, R. Levels of reading diagnosis. Educational Forum, 1969, 33, 2, 187-191.

VanRoekel, B. Personal communication: seminar in diagnosis and remedial reading. Michigan State University, Winter Term, 1978.

Vinsonhaler, J. F. Clinical problem solving among reading specialists: the problem of agreement on a diagnosis. Paper presented at the annual meeting of the International Reading Association, Atlanta, Georgia, April, 1979.

Vinsonhaler, J. F., Wagner, C. C., Elstein, A. S., and Shulman, L. S. Technical manuals for the basic management information system (BMIS). The Institute for Research on Teaching, Michigan State University, 1977.

Vinsonhaler, J. F., Wagner, C. C., and Elstein, A. S. The inquiry theory: an information processing approach to clinical problem solving research and application. Unpublished manuscript, Michigan State University, 1977.

Wade, E. W. The construction and validation of a test for ten teachers skills used in reading instruction, grades 2-5. Unpublished Ph.D. dissertation, Indiana University, 1960.

Wagner, C. C., Kitchen, F., Polin, R., Winsonhaler, J. F., WanRoekel, J., and Gil, D. Technical report: observational studies data analysis system (SDAS). The Institute for Research on Teaching, Michigan State University, 1978.

Weinshank, A. The relationship between diagnosis and remediation in reading: a pilot study (Research Series Number 37). The Institute for Research on Teaching, Michigan State University, 1978.

Wilson, R. M. Diagnostic and remedial reading for classroom and clinic (third edition). Columbus, Ohio: Charles E. Merrill Co., 1977.