The implementation of naturalistic evaluation of reading comprehension: A descriptive study of first graders in a Chapter 1 reading program

Beth Ann Hardy Nordmeyer
University of Northern Iowa

Copyright ©1991 Beth Ann Hardy Nordmeyer
Follow this and additional works at: https://scholarworks.uni.edu/etd

Part of the Elementary Education Commons

Let us know how access to this document benefits you

Recommended Citation
Nordmeyer, Beth Ann Hardy, "The implementation of naturalistic evaluation of reading comprehension: A descriptive study of first graders in a Chapter 1 reading program" (1991). Dissertations and Theses @ UNI. 698.
https://scholarworks.uni.edu/etd/698

This Open Access Thesis is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Dissertations and Theses @ UNI by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.
THE IMPLEMENTATION OF NATURALISTIC EVALUATION OF READING COMPREHENSION:
A DESCRIPTIVE STUDY OF FIRST GRADERS IN A CHAPTER 1 READING PROGRAM

An Abstract of a Thesis Submitted In Partial Fulfillment of the Requirements for the Degree Master of Arts in Education

Beth Ann Hardy Nordmeyer
University of Northern Iowa
August 1991
The purpose of this study was to document the implementation of naturalistic evaluations of reading comprehension, specifically, the comprehension abilities of first-grade students in a Chapter 1 reading program. There were five major questions which guided the study: (a) How will naturalistic evaluations assess students' reading comprehension abilities in a Chapter 1 classroom? (b) What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation? (c) Given special consideration to time and management, how can naturalistic devices be used in a Chapter 1 reading program? (d) How will a Chapter 1 teacher be able to use the information provided by these types of evaluations for improving instruction? (e) Will portfolios provide an evaluation form which is understandable for students, parents, and classroom teachers?

Naturalistic evaluations were found to be effective tools for measuring student comprehension abilities. While running records, retellings, and self-evaluations were useful in yielding sufficient information to be used when planning instruction, observation sheets and journal entries were less effective. Preparing and implementing the evaluations was time consuming as it involved determining appropriate assessments, recording observations, and scoring retellings and running records. However, during the semester, these evaluations became more routine and a natural part of the Chapter 1 program. Parents, students, and teachers responded positively to the
evaluations used in this study. Their comments indicated that they found the information beneficial for understanding student abilities.

Further research documenting the use of naturalistic forms of evaluation is recommended. Practical studies are needed to aid educators in other teaching situations in constructing, implementing, and evaluating student growth through naturalistic evaluations.
THE IMPLEMENTATION OF NATURALISTIC EVALUATION OF READING COMPREHENSION:
A DESCRIPTIVE STUDY OF FIRST GRADERS IN A CHAPTER 1 READING PROGRAM

A Thesis
Submitted
In Partial Fulfillment
of the Requirements for the Degree
Master of Arts in Education

Beth Ann Hardy Nordmeyer
University of Northern Iowa
August 1991
This Study by: Beth Ann Hardy Nordmeyer

Entitled: THE IMPLEMENTATION OF NATURALISTIC EVALUATION OF READING COMPREHENSION: A DESCRIPTIVE STUDY OF FIRST GRADERS IN A CHAPTER 1 READING PROGRAM

has been approved as meeting the thesis requirement for the Degree of Master of Arts in Education

7-8-91
Date

Dr. Jeannie L. Steele, Chair, Thesis Committee

7-8-91
Date

Dr. Charles Dedrick, Thesis Committee

7-8-91
Date

Dr. Junko Yokota Lewis, Thesis Committee

7-24-91
Date

Dr. John W. Somervill, Dean, Graduate College
ACKNOWLEDGEMENTS

A special thank you to the following people who helped me during this writing: Dr. Jeannie Steele for her refreshing feedback and constant reinforcement, Dr. Charles Dedrick for his patient guidance and continual moral support, Dr. Junko Lewis for her inspiration and expertise, my family for their many sacrifices, and the first-grade children for sharing their work and knowledge with me so that I could learn more. This work is dedicated to my husband, Kevin, whose unfltering encouragement was indispensable.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I--INTRODUCTION</td>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Limitations of the Study</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Definition of Terms</td>
<td>3</td>
</tr>
<tr>
<td>II--REVIEW OF LITERATURE</td>
<td>Nature of Reading Comprehension</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Evaluations of Comprehension</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Need for Naturalistic Evaluations</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Naturalistic Evaluation in Chapter 1 Reading Programs</td>
<td>30</td>
</tr>
<tr>
<td>III--DESIGN OF STUDY</td>
<td>Subjects</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Materials</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Procedures and Data Analysis</td>
<td>33</td>
</tr>
<tr>
<td>IV--RESULTS</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>V--CONCLUSION</td>
<td>Summary</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Implications</td>
<td>58</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td>APPENDIXES</td>
<td></td>
<td>63</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Throughout the history of reading research, educational professionals have disagreed about the nature of reading comprehension. Theories of reading comprehension have expanded from placing importance on the process involving rote memory to the process of the acquisition of meaning. Today, educators describe comprehension as the mental act of processing the author's meaning while integrating this information with the reader's schema or prior knowledge. Successful comprehension requires a student to bring meaning to print, not just decode the written symbols (Collins, Brown, & Larkin, 1980; Johnston, 1984; Smith, 1978; Tierney & Pearson, 1983).

Instructional philosophies and practices, such as the whole language philosophy, have kept pace with the evidence provided by research (Altwerger, Edelsky, & Flores, 1987; Goodman, 1987). However, changes in the procedures used to evaluate reading comprehension have not expanded to include the processes involved in comprehension (Joels & Anderson, 1988). It is essential that our evaluation reflect this philosophy, since behaviors which aren't evaluated are not usually taught in the classroom (Farr & Carey, 1986).

Although measures evaluating the process of reading would reflect the accepted view of reading comprehension, product measures are most often used as reading comprehension evaluators. Most of these assessments, including criterion-referenced and standardized tests, measure a student's rote memory which is a product-outcome measure. These norm-referenced
assessments are used to evaluate concepts learned, place students in special programs, and evaluate school programs. Thousands of dollars and hours of instructional time are used to administer standardized tests. Students are currently required to take tests during their academic careers.

Chapter 1 programs have traditionally been remedial; however, with the recent information gained from reading research, these classrooms have become more developmental in their approaches. Chapter 1 instructors teach students the strategies and processes used to construct meaning from the text. These teachers are required to periodically evaluate the students in their class. Evaluations must reflect the philosophy of a classroom (Farr & Carey, 1986). Typically, product-oriented measures have been used to measure a student's growth in a Chapter 1 program. Naturalistic evaluation offers a direct evaluation of the processes a student is using when comprehending (Johnston, 1987a). Considering the changes in philosophy, information about the implementation of naturalistic evaluation in Chapter 1 programs would be beneficial.

**Statement of the Problem**

The purpose of this study is to describe the process of implementing the use of naturalistic evaluation devices, specifically those which measure the growth of reading comprehension, in a Chapter 1 reading program. The following questions will be asked:

1. How will naturalistic evaluations assess students' reading comprehension abilities in a Chapter 1 classroom?
2. What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation?

3. Given special consideration to time and management, how can naturalistic evaluation devices be used in a Chapter 1 reading program?

4. How will a Chapter 1 teacher be able to use the information provided by these types of evaluation to improve instruction?

5. Will portfolio assessment provide an evaluation form which is understandable for students, parents, and classroom teachers?

Limitations of the Study

The reader should remain aware of the following limitations when reading this study:

1. Naturalistic evaluations may include: work samples, anecdotal records, tape recordings, reading conferences, checklists, and retellings (Moore, 1983). The evaluation methods used in this study represent only a sample of those considered to be naturalistic evaluations.

2. The sample population of this research is drawn from first-grade students attending the same elementary school.

3. The students are all in the same Chapter 1 reading program although they have different classroom teachers.

4. The following study is a descriptive of one teacher's attempt to implement naturalistic methods. The teacher is the researcher.
Definition of Terms

The following terms have been defined. These definitions are based on their uses for the purpose of this study.

Anecdotal Record--a brief written description of a behavior, on a note card or paper with the students' names. These observations are reported exactly as observed. The facts of the situation are described and can be used later for interpretation.

Case Study--a type of research which is a comprehensive report about an individual or particular situation. This type of study includes observational techniques.

Chapter 1 Program--a federally-funded program offering additional math and reading instruction. This program is available to schools based on the number of free and price-reduced lunches.

Checklist--a list of reading skills, strategies, or behaviors used by a teacher or researcher when observing a student. This type of a list is used for evaluation or diagnosis, and is later used to improve instruction.

Comprehension Processes--any process a reader uses to acquire meaning. These processes are cognitive in nature.

Conferencing--a discussion between a teacher and student or two students. This interaction centers around classroom happenings, usually writing or reading. Discussions can include a student's past performance, daily work, improvement, and attitudes.

Criterion-referenced Measurement--a measure used to obtain information regarding one's performance compared to the standard one is to attain. Also, this measure compares one student's score to others' scores.
Evaluation—the attempt to understand or judge a student's performance. An evaluation could also judge a person's ability.

Informal Assessments—a non-standardized measure. This measures one's ability or performance.

Naturalistic Evaluation—a nonstandardized form of evaluation used to gain specific information. This information is used to directly improve instruction and the possibility of learning. This form of evaluation includes but is not exclusive to: observations, anecdotal records, journals, self-evaluations, retellings, checklists, and portfolios.

Observations—the act of critically watching students' interactions and signs of growth. Information is gathered, and comments and opinions are made concerning the information gained.

Portfolio—an organized collection of a student's work. A portfolio is used to describe progress of the student's literacy development.

Reading Comprehension—the act of obtaining meaning. Understanding, interpreting or reacting to what was read. The successful accommodation and assimilation of the newly read information with the reader's prior knowledge or preexisting schema.

Retelling—when a person who has read a story describes the events of the story in writing or orally. This can be used as a measure of comprehension and often is included in conjunction with a miscue analysis to evaluate the reader's ability to interpret the text.

Self-Evaluations—occurs when a person critically analyzes their own views, abilities, and attitudes toward learning. Students also may evaluate their progress. Instruments or personal reflection can be used.
Standardized Test—a test with specific tasks and normed scores. The information from this type of a test is used to compare student's measures.

Student Journal—a diary written by the student which includes daily events. Specifically included are events dealing with a student's reading experiences.

Teacher Journal—a diary used to record interactions with students, peers, and administrators. Notes documenting effective lessons and students' growth may also be used for program evaluation and self-evaluation or reflection.

The terms described are significant components of this study. However, in order to understand the current status of reading comprehension instruction, it is important to review its history. Equally as important are the evaluations which have been used to assess reading comprehension. An overview of the teaching and evaluation of reading comprehension is required.
CHAPTER II

REVIEW OF LITERATURE

This review examines four areas related to the evaluation of children's reading comprehension. In the first area, the nature of reading comprehension and instruction are discussed. The second area is concerned with evaluation devices currently being used to measure comprehension with special attention given to both product and process measures. This section also comments on the purposes of evaluating comprehension, the implementation of these evaluations, and the effects of this implementation. The third area focuses on the need to evaluate the complexities of the comprehension process using naturalistic evaluations. The fourth area addresses the use of these evaluations in Chapter 1 programs.

Nature of Reading Comprehension

A major area of controversy in the field of reading is that of reading comprehension. Since the beginning of research in reading, educators have disagreed about the nature of comprehension. A myriad of theories have been developed to aid our understanding of the comprehension process. Robinson, Faraone, Hittleman, and Unruh (1990) summarized educators' views of the reading process throughout history. Prior to 1826, memory was emphasized over comprehension. Throughout the mid-nineteenth century, educators considered reading incomplete without expressive oral reading. During this period of time, decoding was important as a preparation for reading; however, reading for meaning was stressed, and expressive
reading was the ultimate goal. Although, it was agreed that reading was a process, the interlocking and step-by-step models offered contrasting views of reading. The interlocking model suggested teaching comprehension and mechanics simultaneously while the step-by-step process described mastering mechanics before utilizing comprehension activities. At the turn of the century, the thought-getting model deemphasized memorization and offered additional support for thinking, experience, and silent reading. During the twentieth-century, eleven categories of reading methods ranging from exercise and drills (products of comprehension) to schema activation and self-monitoring (processes of comprehension) have emerged for teaching reading comprehension. Goodman (1979) stated that effective reading is defined by successful comprehension. Although, there is no agreement concerning the definition of comprehension, the metacognitive activities used to obtain meaning from a text currently are seen as the final objective of reading instruction (Devine, 1986; Durkin, 1981).

Recent research focusing on metacognition has lead to several conclusions regarding the comprehension process. Valencia and Pearson (1987) addressed the role of prior knowledge and comprehension. They commented that each reader's prior knowledge is diverse and this knowledge varies because of experiences, attitudes, and motivation. If a reader possesses a schema (a pre-existing understanding of the subject) about a topic, when activated, it will aid the reader in comprehending. The mental act of comprehension is also dependent on the situation such as the reader's disposition, the reading atmosphere, and the reader's interest in the topic. Smith (1976) argued that successful reading requires a child to bring
meaning to print, not just decode the written symbols. It has been generally accepted that comprehension is a mental act of successfully processing the author's meaning while integrating this information with the reader's prior knowledge. This prior knowledge affects the way in which a student understands a text. Johnston (1984) studied the effects of prior knowledge on students' abilities to comprehend and concluded, "prior knowledge influences the comprehension of texts and that the effect is not because of contrived materials or other validity problems" (p. 236). From these studies, it can be concluded that prior knowledge affects one's ability or inability to comprehend textual information.

Current research provides evidence that reading involves strategies within the mind of the reader which are dependent upon the text (Collins, Brown, & Larkin, 1980; Tierney & Pearson, 1983). Self-questioning, changing word attack strategies while reading, analyzing, and evaluating the material all happen during the interaction between the reader and the text. These strategies may also include imaging and predicting while reading. These mental acts are known as process-oriented skills. Students who are expert comprehenders vary these strategies depending on the text and the situation (Valencia & Pearson, 1987). Duffy et al. (1987) conducted a study of the effects of teaching mental processing acts as strategies to third grade readers. They reported that students who were taught these strategies became more aware of content and scored higher on all types of reading assessments. Students can be taught a variety of different and effective comprehension strategies that markedly increase their understanding of the
text, often by 50-100% (Wittrock, 1987). These studies are evidence of the importance and potential of using mental processing skills when reading.

As reading theories have expanded from activities involving mainly rote memory and expression to processes of acquiring meaning, so too have instructional practices evolved with reading comprehension research. The whole language philosophy asserts that children learn to read and write naturally, using their own language while integrating the processes of reading, writing, speaking, and listening (Laurin, 1988). The following statements are based on the whole language philosophy. Children learn written language naturally with meaningful and purposeful activities in much the same way they learn oral language. Teachers who maintain the whole language philosophy believe that meaning can not be separated into smaller parts because the whole is greater than the sum of the parts (Altwerger, Edelsky, & Flores, 1987). Therefore, children first learn whole words and sentences not isolated sounds. Skills such as phonics are taught as strategies toward comprehension, not in isolation (Durkin, 1981; Weaver, 1987). The whole language philosophy stresses the process over the products of comprehension using reading, writing, listening, and speaking as interrelated processes (Goodman, 1987; Laurin, 1988). Constructing meaning is stressed over rote memorization. Whole language proponents also believe that children learn best by actively participating in the learning process. Practical applications of whole language differ from traditional skill instruction. There are no absolute methods which accompany the whole language philosophy (Mamchur, 1982). Rather, whole language teachers attempt to teach students to use process-oriented skills by many methods.
Metacognitive processes are encouraged in order to gain meaning. In contrast, traditional classroom teachers who use basal readers do not typically direct their teaching to these mental acts which are so important when attempting to gain meaning.

Watson, Crenshaw, and King (1984) compared whole language and skill-based instruction by observing two teachers whose instruction was based on these two theories. The teachers were taped on eight occasions and 14 people viewed the tapes in order to collect data. These observers concluded that the teacher's actions reflected their theoretical beliefs; therefore, their instruction differed greatly. The whole language teacher focused on large units of language. The skills teacher directed the children's attention to small units of language and considered the exact reproduction of text to denote comprehension. It can be concluded that the skills teacher appeared more concerned with the product of reading rather than the process of reading. The whole language teacher "permitted deviations from the text in that she allowed the children to miscue when reading" (p. 37). In the whole language classroom, the internal process of reading was considered to be more important than the external product. Clearly, the whole language instructor believed acquiring meaning was the main objective of reading.

**Evaluations of Comprehension**

As educators' views toward the nature of comprehension and applied methods have changed, their evaluation procedures have not kept pace. Consequently, these evaluations have continued to reflect the older reading
theories. Moore (1983) presented an historical overview of the measures used to assess reading comprehension. Beginning in the nineteenth century and continuing through the present day, this overview includes recitation, standardized tests, informal measures, criterion-referenced assessments, and naturalistic evaluations. Each of these forms of evaluation will now be discussed.

Recitations

From colonial times to the mid-nineteenth century, evaluation of students' reading achievement was done orally. Memorization and effective public speaking were tested through recitation and oral questioning. This type of assessment provided little if any information about the student's mental comprehension abilities.

Currently, recitation is rarely used as a measure of comprehension abilities, instead assessing comprehension through questioning has been widely used (Sachs, 1984). Durkin's study (1981) provided evidence that basal readers' use of questions to assess reading comprehension is overwhelming. Questions dealt with assessment even when they were labeled as fulfilling another objective. Teachers in this study used questions to assess comprehension rather than to develop comprehension. Durkin concluded that the students learned to provide products such as the right answer and did not learn how to process the meaning of the passages. Duffy and Roehler (1987) also found that because teachers relied on repeated exposure to questions, rather than the mental processes of comprehension for evaluation, students focused on correct answers instead of constructing meaning.
Nessel (1987) described two drawbacks of asking comprehension questions. Many times the types of questions asked require interpretation of relatively unimportant information. Also, the student's ability to answer the question correctly does not necessarily mean that the student is able to comprehend the material being read. Teachers play an important role during assessment. Teachers must ask more than questions which have one correct answer. Nessel commented that teachers should ask questions about what would happen in the story. In order for the students to justify their thinking, Nessel argued that questions must span a range of levels and review the events of the story. However, few commercial materials contain questions which span different levels of thinking.

**Standardized Tests**

Moore (1983) continued his review of evaluation methods by citing Horace Mann's 1845 report in the Massachusetts *Common School Journal*. This report changed educators' views of the evaluation process. Mann argued that printed questions were more objective than oral questions. Mann provided similar situations for those being tested in order for their measures to be compared.

In the early 1900s, the use of test results took on a different focus. Moore wrote:

> It was during this era that many educators became convinced that test results could produce the necessary data on which to base decisions about grouping, grading, referrals, promotions, the curriculum, and the effectiveness of teachers and administrators. . . . Tests were standardized by containing identical tasks, establishing the same testing conditions; and providing uniform criteria for marking and comparing students' responses. (p. 959)
Thus, entered the standardized test. At the beginning of the twentieth century, these tests generally included reproducing short passages and answering questions. "Standardized reading tests allow for comparisons to be made between individual students, classes and schools by establishing standards or norms to which individual test scores may be compared" (Devine, 1986, p. 273). The results of these measures are reported in statistical terms, not always understood by the user. These tests provide relative, general information (Pikulski, 1989). Today, the statistical information from standardized tests is often used to place students in special programs or document student growth (Steele & Meredith, 1991).

Standardized tests are considered to be of some worth because of their consistent procedures and scoring criteria; however, standardized tests have been criticized for their inability to evaluate the interaction between the reader and the text (Devine, 1986). Although standardized tests provide relative, statistical information regarding a pupil's reading ability, they provide little information concerning the processes a student is engaged in while reading. By recognizing and understanding these processes which underlie the student's performance, further instruction could be planned.

Edelsky and Harman (1988) maintain that "reading tests require simulated reading rather than real reading" (p. 158). These authors state that test writers ignore reading as a complex whole and instead propose that meaning moves directly from the author's head to the reader's head. Tests assume that one answer (a product) is an indication of the complex process of reading. Standardized tests provide limited information about the reading process of obtaining meaning. Edelsky and Harman (1988) cite a 1986 study
done by Altwerger and Resta which compared a standardized measure to real reading (oral miscues). Altwerger and Resta found that 1,000 children's California Test of Basic Skills scores indicated no relation to their actual reading performance. They offer the explanation that those who scored higher may have good test-taking strategies rather than better reading abilities.

Regardless of these finding, standardized tests are widely used for the evaluation and diagnosis of many reading areas. Their widespread use results from their easy administration and in some cases, their tested reliability. Standardized measures provide general information about a student’s reading ability. Those who use standardized measures argue that they have traditionally been used; therefore, they will continue to be used (Farr & Carey, 1986).

**Informal Assessments**

In the middle of the twentieth century, informal assessments, such as Informal Reading Inventories, allowed teachers to create tests based on their classes' unique needs. Informal measures may include written and oral retellings and Informal Reading Inventories. Both of these types of measures are currently being used by reading educators.

Story retellings can be used to evaluate a student's ability to recall information from a story. Retellings are valued because they allow the reader to freely recall everything remembered about the text. Retellings can be used as an holistic measure to evaluate a reader's ability to process information in general (Irwin & Mitchell, 1983).
Using story retellings to measure reading comprehension has been the topic of many studies (Kalmbach, 1986). Variations of this technique have also been used. Marshall (1983) stated that the use of a checklist to assess students' retelling performances emphasized the cognitive processes rather than the isolated skills used when reading. Hansen (1978) evaluated students' abilities to retell a story and also answer open-ended questions. Students who were more able to retell a story usually obtained higher scores on the comprehension questions. Irwin and Mitchell (1983) rated retellings on an holistic scale rather than assigning points for each item remembered. They maintained that readers' abilities to generalize the text to their own world are as important as recalling the text's content. Kalmbach (1986) suggested analyzing the structure of retellings as stories. Smith and Jackson (1985) even studied the use of students' writings as the retelling of a story.

An advantage of retellings is that, when retelling a story, students are not forced to adhere to particular questions. It is assumed that, if students are able to recreate the text, then they have the strategies needed to comprehend the text; however, the ability to reproduce material may differ from the ability to gain meaning.

Informal Reading Inventories (IRIs) are typically used to determine students' independent, instructional, and frustration reading levels. Miscue analysis, often done using passages from the IRI, provides information about a student's use of strategies when reading; this allows instructional assistance to be determined (Devine, 1986). Miscues also offer additional evidence about the development of metacognitive skills.
IRIs are considered beneficial because classroom teachers can administer these tests for diagnostic purposes. However, disadvantages have also been identified. IRIs are dependent on the passages they contain and children are asked very structured comprehension questions which may or may not be dependent on the passage. Numerous studies have been conducted to assess the validity of the IRIs' ability to measure comprehension (Duffelmeyer & Duffelmeyer, 1989; Fuchs, Fuchs, & Maxwell, 1988; Joels & Anderson, 1988). Duffelmeyer and Duffelmeyer (1989) concluded that a large percent of Informal Reading Inventory passages, K-6, are not suitable for assessing main idea comprehension. Lehr (1988) instead suggested asking students to identify the theme of the story and then generate statements about the theme.

**Criterion-Referenced Assessments**

Criterion-referenced measures became popular in the 1970s. As opposed to norm-referenced standardized tests, these tests compared a student's performance to a list of specific objectives (Moore, 1983).

Criterion-referenced tests compare students' performances on curriculum-based information to an arbitrary standard of performance. Criterion-referenced measures provide diagnostic information concerning skills not learned instead of providing only grade-level information (Devine, 1986). These tests can measure many specific areas of comprehension, such as main idea, details, cause and effect, and drawing conclusions. However, in order to cover many specific skills, numerous isolated passages need to be read, and mastery of these specific skills does not guarantee that the reader is able to fully comprehend what has been read (Moore, 1983). A criterion-
referenced test assumes that one can break the process of reading for meaning into subparts. Therefore, criterion-referenced tests also have disadvantages.

Naturalistic Evaluations

Naturalistic evaluation is another technique which can be employed to assess children's reading comprehension abilities. Naturalistic evaluations are used to assess the development of the thinking processes and strategies students are using while reading. This form of evaluation has not yet been widely used. After reviewing other forms of evaluation, Moore (1983) concluded that naturalistic assessment is the most promising for useful evaluation. Naturalistic evaluation involves observing children while they are reading. This form of evaluation takes place in students' natural settings. Using the objectives of the class for guidelines, the teacher decides what type of evaluation will be relevant for the individual students.

Goodman (1979) advocated teachers becoming proficient "kid-watchers." Dixon (1987) suggested using teacher observations of children's responses and provides a checklist for assessing oral and written reactions. Some other plausible types of naturalistic evaluations are: observational sheets, video or audio tapes, anecdotal records, samples of the children's work, and portfolios. This type of evaluation is informative, descriptive, ongoing, and determined by the teacher's and students' needs (Moore, 1983).

A teacher must, however, know what behaviors to observe and what implications these behaviors have for children's instruction. Authors have offered additional support for direct evaluation methods developed by teachers which can provide valid information for effective instructional uses
Duffy and Roehler (1987) proposed elaborate responses in which teachers are able to use direct forms of evaluation to discuss mental processes with the students to improve instruction.

Dunkeld and Engle (1980) described an holistic reading and writing project implemented in a six-, seventh-, and eighth-grade Chapter 1 program. In this study, students read and wrote whole pieces of text. Their study included evaluations of the program which consisted of achievement tests, an informal reading inventory, and selected reading passages from trade books. One of the objectives of the project was to observe children's progress and diagnose instructional needs. Clearly, naturalistic evaluations, such as analytic scales, were found to be beneficial when planning instruction. Problems occurred when teachers lacked the knowledge to use analytic scales. Dunkeld and Engle commented on the use of analytic scales which were used to rate the students' writing abilities on a holistic scale:

It appears that the diagnostic information presented by diagnostic tests is different from that provided by the observation of holistic tasks. . . . The experience of the teacher suggests that the valid diagnostic information from holistic activity can be obtained, recorded, and used for instruction when the teacher has been trained in observational techniques related to the process the children are using. . . . Without any firsthand experience of the analytic scales used to evaluate writing or any knowledge of what they contained, the teacher found it difficult to diagnose major needs in written composition and to prioritize instruction. (p. 19)

Portfolios offer an alternative to analytic scales which may be used to aid teachers in establishing and maintaining an evaluative record. Portfolios are used to organize students' work in a usable format (Valencia, 1990).
Portfolios are a collection of the children's work, chosen by the teacher or students. This work could include: writing, self-evaluations, retellings, and responses to books. Each student's work is kept in a separate folder or packet. A portfolio can be used by the teacher to evaluate the children's abilities to comprehend. To make the contents of the portfolio more understandable, a caption is sometimes added. A caption is a piece of paper attached to the work which explains the situation in which the piece was completed, the objectives of the assignment, and whether the student met the objectives. Portfolios are used periodically during a school year, but they may also be collected to show growth over an entire academic career (Valencia, 1990).

Need for Naturalistic Evaluations

Joels and Anderson (1988) concluded that there are disagreements about a reading test's ability to measure comprehension processes; nevertheless, reading comprehension evaluations, especially standardized measures, continue to be used by educators. More importantly, these testing methods continue to influence reading instruction.

Published assessments have a clear and useful purpose for being used to measure reading comprehension; however, as Pikulski (1989) stated, "The value of a test may depend very heavily on the intended use of that test" (p. 81). The assessments described above each have their own advantages and shortcomings. It can be concluded that when choosing a test it is important to remember the purposes and uses of the results. Pikulski also suggests that an evaluation should reflect the type of reading program, goals...
of the instruction being assessed, and the process of reading. Measures should never be used beyond their intended use (Flood & Lapp, 1987).

According to Johnston (1987b), the ultimate goal of educational assessment is to provide optimal learning for all students. Johnston also described three subgoals of pedagogical evaluation as follows: to improve individualized instruction; to select those in need of special programs; and to remain accountable. Each of these goals and their implications will be examined below.

**Improving Instruction**

Information about readers' comprehension strategies is needed for effective instruction. Behaviors which are not tested are usually not addressed in the classroom (Moore, 1983). Leigh (1980) wrote, "Adequate assessment procedures enable the teacher to make appropriate decisions regarding specific program components and to objectively evaluate the results of the program which is developed" (p. 67). Jenkins and Pany (1978) criticized standardized achievement tests for failure to provide information relevant for instruction. When standardized or criterion-references measures are used, task completion, not student understanding, becomes the focus of instructional decisions (Peterson & Clark, 1978).

Published tests have been criticized for their lack of instructionally-useful information. Woodley (1988) provided a classroom frame of reference for teachers to directly evaluate children in their classrooms. He writes:

In many places, American education has come to heavily rely on published tests which yield little, if any, information which can be valuable to the teacher in selecting and planning classroom activities.
Ultimately, the person who must make decisions regarding student learning is the classroom teacher. Teachers need information on an ongoing basis on which to base decisions about the classroom program. (p. 13)

Placement in Special Programs

A nationwide survey conducted by Steele and Meredith (1991) revealed that five standardized tests are most often used for placing students in Chapter 1 programs. These tests are the Woodcock Reading Mastery Test (WRMT), Woodcock-Johnson Psycho-Educational Battery, Wide Range Achievement Test, Peabody Individual Achievement Test, and Brigance Diagnostic Inventory of Basic Skills.

Even though standardized tests are used to aid in the decision of placing students in Chapter 1 programs and to measure growth, the tests' authors did not intend for this to be their purpose. For example, in the Woodcock-Johnson Revised Tests of Achievement (WJ-R): Examiner's Manual (Woodcock & Mather, 1990) the authors' state:

When combined with behavioral observations, work samples, and other pertinent information, results from the WJ-R will assist you in making decisions regarding placement of a student in an appropriate program. . . . WJ-R results may be used to aid in grouping students for instruction within a class or to assemble an appropriate set of students for a special purpose. (p. 8)

The authors also suggest that the WJ-R may be used to assess individual growth and to provide information about program effectiveness, including Chapter 1 programs. Clearly, it is the authors' intention that the scores from this standardized test are not to be used solely for decision making.

Although the standardized test is the assessment measure most often used, it is important to note that Steele and Meredith's nationwide survey also
indicated that the vast majority of states do not rely on standardized tests alone to determine Chapter 1 eligibility. Instead, 79.5% of the States surveyed also use other information along with standardized measures to place students in their Chapter 1 programs. The other criteria which are used include student’s reading level, teacher recommendation, criterion-referenced measures, past grade retention, other test scores, and classroom work.

**Accountability**

It is the need to remain accountable to parents, colleagues, administrators, as well as the general public which has overshadowed the other goals of assessment. Throughout this century, the public has held the scores and percentages produced from standardized tests in very high regard. Testing’s influence is now its greatest in history (Valencia & Pearson, 1987). Teachers must seek ways to gain the public’s trust as well as the trust of administrators and colleagues. It is the educator’s responsibility to provide information to the public regarding evaluation.

Although standardized tests are popular to many, Johnston (1987a) stressed the need for parents to place more faith in the information teachers are able to provide about students’ strengths and growth. Johnston also argues that the same assessment shouldn’t be used for both accountability and instructional planning since a student may react differently knowing that the information is to be used for reporting to parents. If students are worried about the accountability of a test, they will not be as involved in the task. This involvement is needed to provide information for future instruction. Separate evaluations may be used for reporting to parents and
planning instruction or evaluations may need to be done in ways which will be nonthreatening to the students.

The subgoals of pedagogical evaluation are all important to consider when discussing forms of evaluation. However, there are critical ways in which students and the educational system are affected by assessments. Testing has the potential to drive instruction, isolate students, affect students' perceptions of reading, decrease the instructional time used in the classroom, and increase the money spent by a school district.

Test-driven instruction occurs when instructors teach material only because they know that the students will be tested on it. Flood and Lapp (1987) went so far as to recommend this type of instruction. They studied the types of writing in basal readers and in standardized tests, and found that the language used in the basals was very different from that used in the standardized tests. They recommended: "Further, if it is decided that exposition will be given more space in the unit mastery tests and poems and plays less space, then teachers should be informed so that they can make appropriate adjustments in their curriculum" (p. 883). If this is carried out, curriculum would be determined by assessments written by those who have no contact with the students, instead of the teachers who are the ones most qualified to make the curriculum decisions.

Even though educational assessments have been of benefit when identifying students with potential reading problems, they have also aided in isolating some students. Farr and Carey (1986) discuss test bias and its effects. "Data from the broad administration of standardized tests has tended to yield lower scores for urban centered and black populations and
for readers for whom English is not a first language" (p. 47). They continue: "Tests should be selected on the basis of valid information needs that are clearly defined and related to reasonable and accessible instructional goals for a particular population" (p. 47). Owen (1985) cited a 1983 report done by Ramist and Arbeiter in which they found a correlation between SAT test scores and family income. Those with higher SAT scores also reported higher family income. However educational assessments are now at the point where the highly selective enterprise is ending (Glaser, 1988). Glaser states that tests are now focused on the task of creating an entire population of well-educated people. Stayter and Johnston (1990) also express the concern that evaluation can be used in two very different ways: to politically decide a student's future or to improve instruction. Many vital decisions for individual students have been made based on one assessment. "The consequences of such decisions have impact on all aspects of education, and the impact of such decisions on an individual's life is sometimes monumental" (Farr & Carey, 1986, p. 12).

Not only can a student's life be greatly affected by assessments, but their understanding of reading may also be altered because of the assessments. "Assessment practice has the power to influence a child's perception of the task" (Johnston, 1987a, p. 337). Wixson, Bosky, Yochum, and Alvermann (1984) introduced an individual interview procedure used to collect information about a reader's perception of classroom reading activities. This diagnostic information can be used to identify students' misconceptions about reading which may contribute to problems when reading. Students can become confused about the purpose of reading when
given reading tasks which do not reflect the act of reading (Johnston, 1987a). For example, if students are given reading tasks which do not involve meaning, they will perceive decoding as the only purpose of reading.

Johnston (1987a) stated that effective evaluation would include: quality information, information which is likely to be used for instruction, and efficient procedures. Efficiency is particularly important when demands are placed on our instructional institutions to cover more content. The problem occurs when valuable instructional time is used for the administration of tests. Anderson (1982) estimated that during a student's elementary and secondary schools years, he or she may spend two to six hours each year taking published tests. However, this does not include the teacher-made tests and tests which accompany textbooks. It is estimated that in 1987, reading instruction allowed children to read three minutes a day (Johnston, 1987a). The overwhelming cost of standardized tests also needs to be considered when selecting evaluation instruments (Salend, 1984).

Many authors suggest using forms of evaluation which would be continuous, used during instruction, and offer information about students' processing skills (Greene, Doughty, Marquart, Ray, & Roberts, 1988; Ryan, 1985; Studwell & Moxley, 1984; Wittrock, 1987). Each form of comprehension assessment has useful purposes and drawbacks. However, naturalistic evaluations of reading comprehension, especially portfolios, offer many advantages. Valencia (1990) lists these advantages as: (a) naturalistic evaluations are accomplished during instruction;
(b) instructional time is not lost to test-taking; (c) the information gained from these evaluations is authentic, is in the language of the observer, and can be readily used for improvements in instruction; (d) the students are able to become independent learners through self-evaluation; (e) portfolios represent the philosophy that reading is an interactive process; and (f) naturalistic evaluations do not cost a lot of money. Currently used classroom materials are all that is needed. Most teachers already use a form of naturalistic evaluation whether it be an observation sheet or a record of class discussions about what has been read. Naturalistic evaluations are valid and reliable. Moore (1983) writes: "...naturalistic assessment can produce better reliability than traditional approaches because it provides a representative sample of students' behaviors by taking a longer time to gather more observations" (p. 966).

Many aspects of a reader's experience influence the ability to comprehend, yet the reader's attitudes and habits are not always considered as important (Valencia & Pearson, 1987). Most traditional measures of comprehension do not provide information about the attitudes of the reader. If the goal of assessment is to provide optimal learning, as stated earlier, then those doing the testing also need to consider the attributes which may affect students' learning. These experiences may include the habits learned from school or home, the desire to learn or the avoidance from learning because of past experiences, and the ability to accommodate and assimilate new and different information.

Wittrock (1987) discussed process-oriented evaluations which provide information about an individual's comprehension strategies. Process-
oriented naturalistic evaluations provide instructionally-relevant information about individual student's comprehension strategies and thought processes. This information is about individual learners; not group averages. These types of evaluations also provide information about the specific learning strengths and weaknesses of each student on modifiable comprehension processes known through research to be important (Wittrock, 1987). He surveyed 43 reading teachers and found support for this type of assessment. Irwin and Mitchell (1983) described the need for the development and refinement of an assessment device which would reflect the interaction between the text and the reader. From these studies, it is evident that teachers want valid assessment which can be employed in their classrooms and will provide usable information about students' progress.

As methods of teaching reading change, so do the assessments used to measure the skills and strategies learned. In order to provide optimal educational experiences for our students, we must evaluate in a way which is reflective of our teaching philosophies. Steele and Meredith (1991) conclude:

Assessment procedures that accurately reflect the realities of the reading process must include a variety of assessment strategies which reflect this complexity. While the WRMT and the W-J are seen as technically adequate, neither they nor (others) are seen as reflecting recent reading research and as such, do not appear to be assessing what is truly important in the reading process. (p. 18)

If we are teaching children to use strategies instead of skills when reading, it is essential that we have a means of evaluating their development.

Process-oriented behaviors are difficult to assess because they can not be observed directly; therefore, they are usually neglected during formal
Naturalistic types of evaluations are able to provide information about the process of comprehending not just the product, unlike many standardized and criterion-referenced tests. Educational tests typically measure the products of school learning not the processes of acquiring knowledge. For this reason we do not have many techniques which can provide information about how students acquire deep meanings from texts. Because comprehension can not be observed directly, it is necessary to determine an observable behavior which indicates comprehension (Haney, 1985; Henk, 1987; Stiggins, 1985; Valencia & Pearson, 1987; Wittrock, 1987). Johnston (1984) concluded that before an assessment can be developed to analyze a reader's thought processes, a theory and method of generating questions is needed in order for the evaluator to understand the meaning of the reader's response. To date, no such theory has been developed.

Moore (1983) also commented on this dilemma:

The difficulty faced by teachers who wish to refine their naturalistic assessment of reading abilities is finding sources of help. There has been no systematic attention to classroom-based, naturalistic assessment of students' reading comprehension abilities. (p. 966)

Since 1983, there has been more naturalistic evaluation research (Dixon, 1987; Johnston, 1987a; Valencia, 1990; Woodley, 1988); however, there has been no massive swing by practitioners to this assessment approach. This lack of practical application concerning naturalistic assessments of reading comprehension is a serious missing piece in the field of reading.
Naturalistic Evaluation in Chapter 1 Reading Programs

Chapter 1 reading programs offer unique opportunities for students to interact with texts and teachers. Chapter 1 programs have typically been remedial. However, with the changes in philosophy toward the process of reading, these programs have become more developmental in their approach, stressing the cognitive processes of comprehension rather than product outcomes. Because of the intense interaction and small group or one-on-one instruction students receive in these rooms, there are many opportunities and needs for naturalistic evaluation. Yet, Chapter 1 teachers need information about how to implement naturalistic evaluations if they are to gain and use meaningful information regarding their students and their development.

Chapter 1 teachers have a special interest in evaluation because of the federal funding these programs receive. In Iowa, students qualify for Chapter 1 reading services by placing below the 40th percentile rank on a nationally-normed standardized test. Also, in order for schools to maintain their Chapter 1 funding yearly, pre- and post-standardized tests are used to measure the progress of the students in the program. Each district is allowed to predetermine how much growth the students must gain for satisfactory progress (Sabin, 1989). The Chapter 1 teacher may also use other forms of evaluation for classroom use when determining instruction and communicating with parents.

This chapter provides little information about the use of naturalistic evaluations of reading comprehension in Chapter 1 programs. This is because research revealed few studies of this topic. A thorough search was
done using the ERIC data system, and no study specifically documenting the use of naturalistic evaluation in a Chapter 1 reading program was found. Educators need information in order to know what types of behaviors denote comprehension, how to organize naturalistic records, and the best means to develop their abilities as naturalistic evaluators. The study proposed will help to fill this void in the research.

The purpose of this study is to describe and document the implementation of naturalistic evaluation of reading comprehension, specifically portfolios, in a Chapter 1 reading program. This inquiry focuses on the following questions: (a) How will naturalistic evaluations assess students' reading comprehension abilities in a Chapter 1 classroom? (b) What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation? (c) Given special consideration to time and management, how can naturalistic evaluation devices be used in a Chapter 1 reading program? (d) How will a Chapter 1 teacher be able to use the information provided by these types of evaluations for improving instruction? (e) Will portfolios (a form of naturalistic evaluation) provide an evaluation form which is understandable for students, parents, and classroom teachers?
CHAPTER III

DESIGN OF STUDY

This study was designed to document the implementation of naturalistic evaluation of reading comprehension of first graders in a Chapter 1 reading program. This was a descriptive study which used questionnaires, student portfolios, time sheets, and a teacher journal to document information. The study was done during the second semester of the school year and the researcher was the teacher in the Chapter 1 reading classroom.

Permission for the study was obtained from the school's administration and from the student's parents (Appendix A). Furthermore, the research was approved by the Human Subjects Review Board of the University of Northern Iowa. The data which were collected periodically, compiled, and analyzed will be discussed in this chapter.

Subjects

Thirteen first-grade students enrolled in a Chapter 1 program participated in the study. The students were from three different self-contained classrooms of 25 students each. During Chapter 1 instructional time, they were divided into three groups of 4 or 5 students, each representing a self-contained classroom. Each group received 30 minutes of Chapter 1 instruction per day. The students were from a Midwest middle-class, rural community with a population of approximately 2,300 people.
Materials

The students' naturalistic evaluations consisted of the following materials:

1. The students had journals in which they wrote daily.
2. Running records were used to analyze miscues. Marie Clay's method for miscue analysis from *The Early Detection of Reading Difficulties* (1979) was used.
3. The children were asked to retell the story used for the running record. These retellings were scored using a holistic method (Steele, 1990).
4. The teacher used an observation sheet to record instances when students predicted while reading, used prior knowledge, or imagined the story being read.
5. Students periodically conducted self-evaluations.
6. All of the above materials were compiled into folders which were referred to as the students' portfolios.

The teacher employed the following materials:

1. A journal was used to document the daily happenings of the classroom.
2. A time sheet was used to record the evaluations listed above as they were completed.
3. A form (Appendix B) compiling the students' evaluations was used to prepare future lessons.
4. Questionnaires were used to assess the parents', first-grade classroom teachers', and students' understandings of the naturalistic evaluations used in the classroom.
Procedures and Data Analysis

The purpose of this descriptive study was to document the implementation of naturalistic evaluations. In order to answer the questions which guided the study, the materials listed above were gathered and compiled in an organized manner. The following section describes the study questions, materials used to collect information, and the form of data analysis.

How will naturalistic evaluations assess students' reading comprehension abilities in a Chapter 1 classroom? More specifically: What methods are effective? What materials are effective? Why or why not?

The Chapter 1 program's major goal was to develop reading comprehension through the use of strategies. The more detailed objectives were: to anticipate events while reading, to use context clues to analyze words, to stimulate previous knowledge, to imagine the story while reading, to develop an understanding of story elements including the events, characterization, and setting, and to become aware of when these strategies are being used. Each of these objectives and the ways in which they were evaluated will now be discussed.

Observational sheets (Appendix C) were used every other week to record when students were predicting while reading, stimulating previous knowledge, or imagining the story. These formal observations occurred during an entire week whenever the children were reading alone or with the group. There was an observation sheet for each student. This sheet identified the behaviors being observed: predicting, commenting on previous experiences similar to that in the story, or describing the story as
the student imagines it in his or her mind. The teacher marked the number of times the behavior was observed on a line next to the listed behavior. At the end of the week, the teacher tallied the number of incidences in which each behavior was observed. These were interpreted as follows: 0 times: no indication of the strategy; 1-2 times: the strategy was beginning; 3 times: the strategy was developing. If the strategy was observed more than three times, it was considered to be an independent strategy.

Self-evaluations (Appendix D) assessed whether the students were aware of when they were using the comprehension strategies: predicting while reading, using context clues, or imagining the story. It was also possible for the teacher to investigate whether the student's perceptions reflected their abilities. The students were asked to what degree they agreed or disagreed with statements concerning their abilities. For example, they were asked whether they predicted while reading and they were asked to draw a line to the word yes or the word no. They could draw a short or long line; the longer the line was drawn toward these answers, the stronger they agreed or disagreed. This allowed the students to respond very specifically. The self-evaluations were completed once at the end of every month (January through May).

Every other week, the teacher compiled all the evaluations used (Appendix B). Comparisons were made between the students' perceptions and the observations of the teacher. Any comments regarding these comparisons were noted on the compiled evaluation sheet.

Running Records were used to analyze the word attack strategies the children were using when reading. Marie Clay's running record procedures
(Clay, 1979) were used (Appendix E). The miscues were recorded on a running record (Appendix F) and analyzed. The children's retellings were recorded onto a cassette tape and given an holistic rating (Appendix G). Then the percentages of miscues, percentages of context matches, and the holistic retelling scores were recorded onto a sheet where the students' word attack strategies could be compared to their abilities to retell the story (Appendix H). A running record and retelling was done every other week, alternating with the observations.

The students' journals were used to evaluate all of the objectives depending on what the child had written. Every time the students wrote in their journals (once a week), the teacher recorded the journal content on each child's note card. This was done because the students may have been the only ones able to read the journal content. If the information was an indication of one of the objectives of the program, it was added to the observation sheet, running record and retelling summary, or self-evaluation sheet.

At the end of the semester, each student's evaluations were collected on one summary sheet and analyzed (Appendix I). The following questions were used: Did the student show growth with any form of evaluation? Did the self-evaluations change?

What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation? More specifically: How will they be created and organized? Will these evaluations be efficient? Where will the evaluations be kept? How much additional time is needed to interpret the evaluations?
A teacher journal was used to record the problems which were encountered when the naturalistic forms of evaluation were implemented during the semester. The teacher wrote in the journal as frequently as possible. At the end of the semester, the teacher divided the comments by similar topic and examined them for consistent or important problems.

Given special consideration to time and management, how can naturalistic evaluation devices be used in a Chapter 1 reading program? More specifically: How will the teacher deal with only 30 minute periods of instruction? Will this be enough time to listen to 13 students read, retell, or self-evaluate?

Every other week, the teacher filled out a time sheet (Appendix J). This incorporated all of the evaluations done during the proceeding two week period, whether or not they were completed, and why. At the end of the semester, the teacher tabulated the number of evaluations completed compared to those not completed. Also, the reasons why the evaluations were not completed were tallied so that the teacher could look for a consistent problem.

How will a Chapter 1 teacher be able to use the information provided by these types of evaluations to improve instruction? More specifically: How will the evaluations be compiled and organized for later use? Are these evaluations helpful? Is the teacher able to determine what instruction would be beneficial?

The teacher used the compiled evaluations sheet (Appendix B) to look for the students' instructional needs. At the end of the semester, the teacher analyzed the compiled evaluations sheet and asked: Was the sheet utilized?
The teacher also asked: Was the instructional need of each child clear? The teacher commented in the teacher journal concerning these questions.

**Will portfolio assessment provide an evaluation form which is understandable for students, parents, and classroom teachers?** More specifically: Do they understand the evaluations being used? Do the teachers use the information provided by the evaluations differently than they would information from other forms of assessment?

At the conclusion of the semester, classroom teachers and parents were given a questionnaire and the students were interviewed (Appendix K). The information gained from each group was totaled and compared.
CHAPTER IV

RESULTS

The purpose of this study was to document the implementation of naturalistic evaluations of reading comprehension in a Chapter 1 reading program of first-grade students. Five major questions guided the study. The results concerning each study question are presented in this chapter.

How will naturalistic evaluations assess students' reading comprehension abilities in a Chapter 1 classroom? More specifically: What methods are effective? What materials are effective? Why or why not?

The methods used varied in their effectiveness. While running records, retellings, and self-evaluations were useful in yielding sufficient information concerning children's comprehension abilities, observation sheets and journal entries were less descriptive. The researcher measured the evaluations' effectiveness based on the cumulative evaluation sheet on which each student's performance was summarized. It should be noted, however, that growth was not necessary to determine an evaluations' effectiveness. Rather, the researcher studied the summary sheet to uncover whether or not there were any changes or consistencies in the students' comprehension abilities. Each of the specific evaluations will now be discussed.

Running records and retellings were helpful when determining growth of comprehension skills. All students displayed varying abilities. As could be expected, the students' comprehension skills were dependent on their interests in the topic being read. Most valuable was the percentage of
miscues which made sense in context. The teacher was able to compare the number of miscues making sense in context with the number of total miscues. The students' abilities to use context was compared to the holistic retelling score. This information was similar in most instances. For example, Devin read one story with 88% accuracy; 8 of the 15 errors made sense in context. If the miscues which matched context were counted as correct, his reading accuracy was raised to 94%. The retelling of this story was rated a 3 on a scale of 0 to 3 with 3 being highest. When the information was not similar, it could be compared specifically to determine why the differences might have occurred.

The students' self-evaluations were very effective. Each child's self-evaluations were compiled onto one sheet in order to compare how they changed during the semester. All the children made changes in the evaluations from month to month. Most of the time these changes matched the growth that the children were experiencing in the other evaluations. For example, during the week of March 25, Josh commented on his self-evaluation sheet that he did not make predictions. During the same time, he was observed making predictions only once. Later, during the week of April 29, Josh evaluated himself again and said he was now making predictions. He was then observed making predictions three times during a one-week period. There were instances, for example, when a child would comment that he did imagine when reading, the next month he would comment that he did not, and the following month say again that he did imagine. When these inconsistencies occurred, it was beneficial to add the comments the students made about their self-evaluations while completing them. For
example, one student commented, "I didn't know what that meant before. Now, I get it." Some students even wanted to go back and change a previous self-evaluation because they now knew that what they said earlier was not what they really felt at the time.

Observation sheets were somewhat less effective than the retellings and running records. It is difficult to compare the observations taken at different times because there were other factors entering into the situation. For example, the number of times "imagining" was observed is questionable. Perhaps on one day, the students were willing to share the images they had of the story. The teacher then recorded these comments on the observation sheet. On another day, the students might not have commented on the pictures they imagined. This change could have been because they had no images, because they did not feel inclined to share their ideas, or because they felt intimidated by the situation. A zero or low score on the observation sheet did not necessarily mean the students were not imagining. For this reason, the observation sheet's validity was weakened and the value of the assessment was decreased. During the week that the observations were taken, it was important to obtain a broad sampling of information from more than one class period. This was necessary in order to ensure that the changes in behavior were sincere and not because of another reason, such as those listed above.

The student journals were least effective of all the evaluations. Because their journals were viewed as their private place to write information of their choice, the students rarely found it appropriate to comment in their journals about their reading abilities. It was intended that
the students would write in their journals regarding their abilities only when they found it appropriate. It was also intended that comprehension type materials would be kept in the journals only if the need arose, which did not occur. Therefore, there was little information in the journals to use to measure their reading comprehension. Most of journal time was spent writing daily happenings. On one occasion, the students did make one or two comments each concerning their reading abilities such as, "I no a lt w" (I know a lot of words.) and "I ON HW to u br the pr" (I know how to use the pictures.). However, this was only because they were asked to do so by the teacher.

The summary sheet was extremely beneficial when determining the students' strengths and weaknesses. Not every child showed growth, but there were consistent patterns for each child which could be explained after looking at the evaluations all listed together. It would have been very difficult to find general comprehension strengths and weaknesses without the use of the summary sheet. This sheet especially helped to identify changes in behavior which happened during the same time because the evaluations were listed chronologically. The running record and retelling could be compared to the observation sheet and self-evaluation done at the same time (see Appendix L).

What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation? More specifically: How will they be created and organized? Where will the evaluations be kept? Will these evaluations be efficient? How much additional time is needed to interpret the evaluations?
Preparing the evaluations to be used was time consuming. It took approximately 20 hours to determine the program's goals, write the evaluations, and prepare these evaluations to be used. After the evaluations were written, it took additional time to revise them to be used. The teacher created the evaluations by listing the comprehension goals of the program. All possible naturalistic forms of evaluating these goals were listed under each goal. It was important to review all the possible naturalistic forms of evaluation in order to choose the ones most appropriate for first-grade students.

All necessary sheets of paper were photocopied prior to the beginning of the study and kept in file folders which were labeled. This included the self-evaluation sheets, the observational sheets, and the running record sheets. The journals were constructed with a large piece of construction paper folded over writing paper. The construction paper served as the front and back covers. The students' names were written on the front of each of their portfolios and the cover of their journals. The portfolios were colored folders purchased by the teacher. Each of the students in a group (the first graders represented three separate classrooms and came in groups divided by these classrooms) had the same color folder. This made it easier to organize the portfolios without having to take time to read the students' names.

The portfolios were kept in a stackable divider on a table in the room. They were kept where the children were able to reach them. The journals were stored in the portfolios. When the students filled a journal, they took that journal home and another journal was constructed. At first, the
observation sheets were kept in a three-ring binder. The teacher found this very difficult to maneuver, so the observation sheets were held on a clipboard during the observational days. The cards used to record the students' journal entries were taped onto a clipboard. The running record sheets were also clipped on the clipboard until they were evaluated. This helped the teacher remember to score them. The observation sheets were kept clipped together by class until all the observations were recorded. The running record and retelling sheets were also kept together by class until they were scored. Then all the evaluations were put into each student's portfolio.

The evaluations were not efficient at the beginning of the study. The teacher and students needed time to become familiar with the evaluations. At times, students would comment on something which the teacher felt was beneficial or representative of their comprehension. This might have happened when the observation sheet was not with the teacher or when more than one child was commenting at a time. If the teacher did not write the comment down immediately on the observation sheet, the information was lost forever. It is estimated that this probably happened at least three to five times during the study. It was also difficult to obtain a running record without an interruption from one of the other students.

During the second month of the study, the evaluations became routine. The observations were done during regular class time when the students were commenting on their reading processes. The teacher would comment on what the student had said and then mark the observation sheet. The conversation was rarely interrupted. When it was, it was because the
teacher wanted to write down what the student had said, verbatim. The running records also became routine. As each student read to the teacher, the teacher was able to mark the running record sheet quickly. The other students learned not to interrupt, allowing the process to become faster.

The students took an active role in recording the evaluations. The retellings were tape recorded by the students. This allowed the students to be working on other activities with the teacher until it was their turn to retell the story. It was sometimes possible for one child to be retelling a story into the tape recorder while another read a story for the miscue analysis. When this was done, the other children wrote in their journals. Then the students rotated work areas. The teacher made an effort to listen to the retellings on the same day they were taped. When this was not possible, the tape was saved until a later time. While listening to the retelling, the teacher took notes. These notes were then used to assign a holistic score to the retelling (Appendix G). In order to obtain a comparable measure, it was important that the retellings be done on or near the same day as the running records.

The time necessary for compiling the evaluations depended on the type of evaluation. Counting the miscues on the running records was the most time consuming. It was necessary to allot specific time to be used for analyzing the running records and retelling. It took approximately 20 minutes to listen to a student's retelling, assign it an holistic score, count the miscues on the running record, also count the number of miscues which made sense in context, and record all the information on the cumulative sheet so it could be examined for similarities. With 15 students, this took 4
hours and 20 minutes every other week when the students were evaluated. Extra time was not necessary when totaling the observation sheets. This could be done for the whole class during the five minutes at the end of the class period as the students were putting their materials away. A classes' self-evaluations could also be recorded on the compiled evaluations sheet in approximately five minutes following the class period.

A problem which developed concerned the time needed to share the information with the teachers. Since the teachers all arrived early, this was usually done before school on a drop-in basis. Written notes were also used when necessary. The teachers were asked to respond in writing to the evaluations, but none did so. Consequently, these notes provided the information for the first-grade teacher but did not provide for any information to be returned to the Chapter 1 teacher.

The word portfolio was sometimes a problem. It was a difficult word for the students to pronounce. One first-grade teacher suggested that it be called a folder, since folder was easier to remember and pronounce. But, the word portfolio was used daily in the Chapter 1 room so it did become routine and easy to remember. Portfolio carried a more special meaning and importance than the word folder. This special name helped the students understand that it was a place to keep very important individual work.

Given special consideration to time and management, how can naturalistic evaluation devices be used in a Chapter 1 reading program? More specifically: How will a teacher deal with only 30 minute periods of instruction? Will this be enough time to listen to 13 students read, retell, or self-evaluate?
Not all of the evaluations were completed. If the running records and retellings were to be done every other week, seven should have been completed during the study. All but two were completed. In both instances, the running records and retellings were not done because the students were busy completing writing activities or participating in a school-wide event which interfered with the Chapter 1 meeting time. The class should also have completed eight observation sheets. All the classes missed two of these because situations were not available to observe them imagining, using prior knowledge, or predicting. All of the self-evaluations were done, although two individual children missed one assessment each and this was because they were absent. When this happened, the assessment was done at another time. All of the weekly journal writing was completed.

One of the main reasons for not doing an evaluation was because the students were out of the Chapter 1 room, not because time could not be found within the class period. A unique problem to the Chapter 1 setting, was that other special activities sometimes overlapped the reading class time. If students are scheduled to be in a Chapter 1 room and other school-wide activities are taking place, the school-wide activity had precedence. It was rarely possible to reschedule meeting times since the Chapter 1 schedule was full already.

Marie Clay's method of taking a running record was very practical. This method did not require the teacher to have a copy of the text. This made it possible for the teacher to obtain a miscue record of any text at any time. Nevertheless, it was important to review the original text in order to count miscues which made sense in the context. A blank sheet of paper was
used with space for the students' name and title of the book at the top (Appendix F). However, in order to save paper, the teacher often used scratch paper which had been printed on the reverse side. Running records only took one side of the paper when the teacher wrote in two columns (Appendix M). The teacher usually scored the miscues after school on the same day as taking the running record. If time was not available to do this on the same day, the teacher put the sheet of paper containing the running record in the book and kept it until time was available. The score was written on the reverse side of the running record sheet. A percentage of correctly read words was recorded as well as the percentage of miscues which made sense in context.

At first, the assessments took a longer period of time to complete. At the beginning, the self-evaluations took an entire class period. Later when the students were familiar with the format, it took 10 to 15 minutes for each to complete the form. The running records also took a longer period of time at the beginning. For example, it would take two class periods to listen to everyone read. As the study progressed and everyone became familiar with the procedures, however, it only took one class period to listen to everyone in the class (four or five students) read the story. Observations were not difficult because it was planned to use an entire week to record information. It took this much time to record a reflective sampling of the children's abilities.

How will a Chapter 1 teacher be able to use the information provided by these types of evaluations to improve instruction? More specifically: How will the evaluations be compiled and organized for later use? Are these
evaluations helpful? Is the teacher able to determine what instruction would be beneficial?

The evaluations were compiled on one sheet per class every two weeks (Appendix B). This form was organized by each child's name and the evaluations done. The forms were kept in a three-ring binder on the teacher's desk. The teacher copied the evaluation results from the other forms (Appendixes C, D, and H) at the end of the two week period. After compiling the information, the teacher was able to determine consistent patterns for each child.

This form was especially helpful, however, because the teacher was able to identify similarities among students from the same class. This aided in planning instruction for the entire class. For example, during the two-week period of April 29 to May 6, it was noticed that one group was overwhelmingly relying on phonics or sight words and not using context to decode words. Consequently, this group benefited from instruction encouraging the use of context. Also, the group contained two students who were developing their abilities to predict more than the other three students. This class also benefited by the two boys sharing their predictions with the class.

Every time the teacher wrote down information concerning the children's progress, the previously recorded information was noted. This allowed the teacher to be reminded of the child's growth. This also helped to gain an understanding of what type of instruction was not effective previously. The teacher did make an attempt to address each of the needs of the children.
Will portfolio assessments provide an evaluation form which is understandable for students, parents, and classroom teachers? More specifically: Do they understand the evaluations being used? Do the teachers use the information provided by the evaluations differently than they would information from other forms of assessment?

The students were interviewed and asked, "Do you know what is in your portfolio? Describe it to me, telling me what it tells you."

To the first question, the students replied one or more of the following: "My journal, self-evaluations, papers, work, or report card." When asked to describe what it tells them, one had no comment. Of the remaining 12 students, four had short comments such as, "... to write better ... to learn more ... tells us what we know ... so we could remember what we've learned." Eight of the students made longer comments. These comments could be divided into groups. Some of the comments centered around the specific evaluations being used: "... self-evaluations help us see how much we've learned ... we self-evaluate so we know if we changed ... journals remind you of how bad you wrote ... journals help you remember what you wrote." Other comments considered the past while looking toward the future:

... it shows how well you read in the past few years ... tells us what we do ... you write stuff so you can remember what we have worked on ... if you didn't know something before, than we can learn it ... so we see one thing you might change the next time ... because if you do the same thing all the time, you'll be bored ... we work on what we need ... so we can save all the stuff so we can see how we did at the end of the year ... compare it to what we did before.
One student even contemplated the role of his parents when he said, "... so we can see what's wrong ... our moms can read it to us and know how to help us." It is apparent that the students understood the use and benefits of the portfolio.

The students took an active role with the evaluations, for example, when tape recording the retellings. This appeared to help them to become more aware of their weaknesses and how they could improve upon their strengths. At first they did not understand how to self-evaluate. It was necessary for the teacher to model the thinking involved.

The parents' responses were very positive even though only six parents returned the questionnaire. The parents were asked to rate the following evaluations: self-evaluations, observations, portfolio contents, and journals. They were asked to rate them in descending order as very beneficial, helpful, "did not matter to me," or not beneficial. They were also asked to comment if they did not understand the information. On the questionnaire, all but one of the responses indicated that the parents found the information very beneficial or helpful. More specifically, two found the students' self-evaluations very beneficial, four found them to be helpful. Five parents thought the teacher's observations regarding their child's reading was very beneficial while one found the information to be helpful. One parent commented that the information in the child's portfolio was very beneficial and four thought it was helpful. Only one person indicated that they did not understand the information in the child's portfolio. This person wrote, "I understood when she explained it to me, but just to look at it I had no idea what it meant." Five persons said the information in the child's
journal was helpful and one said it was very beneficial. Even though less than 50% of the parents returned the questionnaire, the response was overwhelmingly positive and supportive of naturalistic evaluations.

One parent wrote extensive comments concerning her child's evaluations. She wrote, "I thought this was interesting because it showed how the child felt about his own abilities. In Bryce's case, by third quarter he knew he was learning to read and felt good about himself." When commenting on the teacher's observations, she wrote, "I feel this is the most helpful. It helped me to understand what they were trying to accomplish and how I could help." And finally, she wrote the following about the portfolio in general, "The portfolio helped me to understand what they did and how they were evaluated. It also showed how much they improved as time went on."

Some parents also made comments during conferences. One father said, "This way it seems you know a lot about each kid to be able to write everything they can and can't do." This parent was referring to the quarter evaluation which were completed. This quarter evaluation included the quarter's objectives, the child's progress, and the goals for the future. The information on the compiled evaluation sheet was used to do this. Another parent felt that her child was gaining more from the evaluations than just an understanding of what he could or could not do. She said, "When Josh sees how he did on a test, he puts himself down and then he loses interest in reading."

The three teachers had mixed views on their questionnaires. The teachers were asked to rate the following evaluations: self-evaluations,
observations, portfolio contents, and journals. They were asked to rate them in descending order as very beneficial, helpful, "did not matter to me," or not beneficial. They were also asked to comment if they did not understand the evaluation. They all rated the information found in the evaluations very beneficial or helpful. One teacher stated that the comments helped her to understand her students better. Another teacher was very concerned that parents wouldn't understand that the Chapter 1 assessments were done in a small group setting and that she observed other behaviors in the classroom. She went so far as to say that the Chapter 1 evaluations made her "look bad."

The teachers were also asked to rate the evaluations based on how they used the evaluations for their own instruction. In descending order, the ratings were extremely useful, quite useful, moderately useful, of little use, and of no use. Two of the teachers felt that the observations were extremely useful when planning their own instruction. One said that the observations were of moderate use. One teacher thought the portfolios were extremely useful while two thought they were quite useful. The teachers also found the self-evaluations and journal entries moderately useful or quite useful.
CHAPTER V

CONCLUSION

The summary, discussion, and implications of this study are presented in this chapter. The chapter begins with a brief summary of the first four chapters of this report. Next, the findings of this study are discussed. Finally, implications for creating naturalistic forms of evaluation and further research are given.

Summary

The professional literature describes the many views concerning reading comprehension (Robinson, Faraone, Hittleman, & Unruh, 1990). These varied opinions have existed since the beginning of reading instruction itself. During the twentieth century, methods ranging from drill to schema activation have been used for teaching comprehension. Recently, reading comprehension has been described as one involving metacognitive processes (Devine, 1986). This view is shared by many reading researchers. Successful reading requires a student to bring meaning to the print (Smith, 1976). The student's prior knowledge affects comprehension (Johnston, 1984). Students who are able to comprehend vary their process-oriented skills such as self-questioning, imaging, and making predictions (Valencia & Pearson, 1987). The whole language philosophy stresses the mental processes used when comprehending (Goodman, 1987).

The literature also includes a variety of assessments used to measure reading comprehension. As the views of reading have changed, so too have the evaluations used to measure reading comprehension. However, changes
in the understanding of reading processes have advanced beyond assessment changes. Questioning, standardized tests, informal assessments, and criterion-referenced assessments, have all been used to measure students' reading comprehension. All these forms of evaluations have their benefits. However, considering the metacognitive, processing view of the reading process, these forms of assessment all have disadvantages as well (Duffelmeyer & Duffelmeyer, 1989; Edelsky & Harman, 1988; Moore, 1983; Nessel, 1987). One of the problems is that these types of evaluations have not expanded to include metacognitive processes (Joels & Anderson, 1988).

Naturalistic forms of evaluations have been recommended to measure a student's comprehension abilities (Moore, 1983; Valencia, 1990). Their advantages range from providing useful information for improving instruction to reflecting the actual interaction between the text and the reader (Irwin & Mitchell, 1983; Wittrock, 1987). However, a noticeable void of information exists concerning the use of naturalistic evaluations. In particular, there is a lack of information regarding the use of naturalistic evaluations in a Chapter 1 setting.

This descriptive study attempted to investigate how a teacher would implement naturalistic evaluations in a Chapter 1 program and the problems which would be encountered. An effort was made to document the steps which were taken to organize and create the materials and the time which was spent recording the assessments. The researcher also took into consideration the views of the parents, classroom teachers, and students, and their reactions to the assessments. Five major questions guided the study: (a) How will naturalistic evaluations assess students' reading comprehension
abilities in a Chapter 1 classroom? (b) What problems does a Chapter 1 teacher encounter when implementing naturalistic forms of evaluation? (c) Given special consideration to time and management, how can naturalistic evaluation devices be used in a Chapter 1 program? (d) How will a Chapter 1 teacher be able to use the information provided by these types of evaluations to improve instruction? (e) Will portfolio assessment provide an evaluation form which is understandable for students, parents, and other teachers?

The class selected for the study consisted of 13 first-grade students, 12 of whom were boys. They were chosen to participate in the Chapter 1 program based on a standardized assessment as required by the federal government and teacher recommendation. The 13 students were divided into three groups based on their classroom assignments. Each group received 30 minutes of instruction in the Chapter 1 reading room.

Because of the scope of the study, it was necessary to limit the number of naturalistic evaluations which were used to measure reading comprehension. The following evaluations were used: journal entries, running records, retellings, observation sheets, and self-evaluations. The teacher, who was also the researcher, developed and employed time sheets, a compiling sheet to record and compare all a student's evaluations, as well as questionnaires for parents, classroom teachers, and the students to monitor the use of the evaluations.

The evaluations varied with their effectiveness to assess literacy development. Running records, retellings, and self-evaluations were effective in documenting useful information concerning students'
comprehension abilities. Consistencies were found between these assessments' measures. Less effective were observation sheets, used to record prediction making, use of prior knowledge, and use of imagery. It was difficult to compare the observations to the other forms of evaluation because there were other variables affecting the observations. Students were not always willing to share what they were imagining or predicting so it is doubtful whether all the observations were reliable measures of comprehension. The students' journals were also less effective, but for a different reason. The students were allowed to write their own ideas in their journals. This content rarely contained any information concerning comprehension. Therefore, the information in the students' journals was of little use for measuring comprehension.

Much time was spent preparing and recording the forms of evaluation. The materials were organized in the classroom for ease of use. The teacher kept all materials until the results were determined and the information compiled. Then the information was put in each child's folder, known to the students as their portfolios. At first, the evaluations were time consuming and cumbersome. As time passed during the study and the teacher and students became more familiar with the evaluations, the evaluations became more efficient and routine.

A few evaluations were not completed. This was because the students were absent, there were other activities which needed to be completed, or the students were taking part in a school-wide activity and did not attend the Chapter 1 class. An effort was made to "make-up" any assessments
which were missed; however, when a whole class missed an evaluation this was very difficult.

The evaluations were compiled onto one sheet where comparisons could be made and patterns of behavior could be found. These compiled sheets aided in identifying common problems among students from one class so appropriate instruction could be determined. These sheets also made it possible to check if a student demonstrated similar comprehension abilities based on different evaluations.

As a whole, the parents, teachers, and students found the naturalistic evaluations to be useful and understandable. Only six parents returned their questionnaire; however, their comments were positive and supportive of different types of evaluation. The students indicated that they understood what they were able to do and what they needed to learn. The teachers also found the materials to be helpful. Two of them commented that the information gained by the evaluations were extremely useful when planning their classroom instruction.

Discussion

Enough information was gained to indicate that naturalistic forms of evaluation can be used to measure reading comprehension. The Chapter 1 program offers a unique setting in which to evaluate as well as provide needed instruction to reading students. The information from the assessments used in this study were shown to be understandable to those involved. Furthermore, they were useful in the improvement of instruction. Based on this study, naturalistic assessment can be useful in other Chapter 1
settings. It is clear that modifications of the evaluative process would create an even more effective evaluating situation and provide more useful information to be used for improved reading comprehension instruction.

It is recommended that observation sheets include a longer period of time (two to three weeks). This researcher took observations for the duration of one week. A longer period of time is necessary in order to obtain a realistic sampling from the students and eliminate any other contributing factors which may keep the students from sharing their predictions and what they were imagining. For example, when the students did not share during this study, it was assumed they had nothing to share and consequently important information may have been lost because it was a short period of time and not a large sampling.

It is also recommended that journal entries not be included as a formal part of the evaluation process for reading comprehension. Little information was gained from the journal entries. This information could be added to the cumulative sheet later if it appeared to be relevant. However, much time was spent recording the journal entries and little information concerning the students' comprehension abilities was obtained.

Since extra time was needed in order to compile the information gained by the retellings and running records, it would be more realistic to obtain these measures every month instead of every two weeks. This study was done during one semester and seven retellings and running records were done. It was possible to show growth based on these seven evaluations. Considering that a school year lasts nine months, it should be
possible to find growth over a period of seven to nine months with one
retesting and running record evaluation done each month.

It was difficult to find time to meet with the students' classroom
teachers. Meetings with parents and other teachers did conflict with
scheduling meetings with the classroom teachers. It would be more efficient
to have a permanent meeting time every other week or every month. This
would help ensure that the classroom teacher could meet with the Chapter 1
teacher for uninterrupted conference time.

Another problem concerned the scheduling of the Chapter 1 classes.
These classes were scheduled one right after another. It would have been
helpful to have a five or ten minute delay between the classes. The classes
allowed enough time to assess with the student, but the teacher would have
benefited from an additional five minutes before the next class arrived in
order to add any comments to evaluations after the students left.

Even though the evaluations were not efficient at the beginning of the
study, it is doubtful that this process can be avoided. It is necessary that
both the teacher and students become accustomed to new ways of
evaluating. This is a normal, expected part of implementing any new
program. This is especially true for a program which involves the class in an
active way, such as using observations and self-evaluations. It is important,
that teachers using naturalistic assessments which they have written reflect
on their own learning experiences in order to improve from year to year.
This is necessary to increase efficiency, effectiveness, and usefulness of the
assessment program.
Implications

Reading comprehension will continue to be a major goal of all reading programs. As long as comprehension is a main objective of reading instruction, teachers should continue to seek effective ways to evaluate comprehension. This information is crucial if teachers are to provide well reasoned, effective instruction. Given the information concerning one teacher's implementation of naturalistic evaluations which is provided above, teachers are encouraged to experiment with these and other forms of evaluation of reading comprehension.

This study filled only a small part of the vast void which remains to be filled concerning the use of naturalistic evaluations. The use of only a few forms of naturalistic evaluation has been documented here. The investigation provides information concerning the construction of naturalistic forms, the implementation of the assessments, and the difficulties experienced while the assessments were used. However, further research of naturalistic forms of assessment is necessary. If assessments are to reflect the specific situation in which they're implemented, practical studies such as this are needed to aid teachers in other teaching situations in constructing, implementing, and evaluating student growth using naturalistic assessments.

In conclusion, researchers are encouraged to create their own naturalistic assessment materials or use those already constructed by others. They are encouraged to document their use so that others may learn from their successes and failures. As our views concerning the teaching of reading grow and change based on the what we understand about the learning process, so too must the way we assess. The area of assessment of reading
deserves further attention. If our evaluations do not keep pace with our methods of teaching, we risk resorting to less effective teaching methods which can be easily implemented, yet do not accurately reflect the reading process. If this happens, the knowledge we have gained concerning the reading process will have been in vain and assessments which do not reflect this process will determine our instruction. Students deserve the most effective instruction and evaluation methods we can provide. Naturalistic evaluations offer an effective, practical, and understandable alternative for assessing reading comprehension in the classroom. The use of naturalistic evaluations in classrooms, especially Chapter 1 reading programs, deserves further attention.
REFERENCES


Sabin, B. J. (1989). The conflict between instruction that characterizes Chapter 1 programs and instruction that results from the whole language philosophy. Unpublished research paper, University of Northern Iowa, Cedar Falls.


Appendix A

Dear Parents:

November 12, 1990

In order to understand your child's needs in reading, I will be evaluating your child's progress throughout the year. Each day I will be listening to him/her read and I will observe him/her while reacting to the material. Your child will evaluate his/her own work by commenting in a daily journal and keeping daily work in a portfolio. This is a collection of work which is kept to show the growth throughout the year. Each child will choose what is included in his/her portfolio. The types of assessment described are called naturalistic evaluations because they reflect natural, real reading which occurs in a typical classroom.

As part of a graduate research study at the University of Northern Iowa, I will be documenting the use of reading evaluations. The purpose of the study is to provide information to teachers about ways to assess reading comprehension. I am also interested in how parents feel about different forms of evaluation. This type of information is vital in order in improve students' educational experiences. During the 1990-91 school year, I will be keeping a record of how naturalistic evaluations are used in my Chapter 1 Reading room. The evaluations are part of a normal Chapter 1 experience only I will be recording how these assessments are managed, what problems may be encountered, and if parents and other teachers find the information useful.

I am asking for your permission to include information from the evaluations used with your child. Your child will not be identified, all
information will be kept confidential, and all participation will be voluntary. Refusal to participate will result in no penalty. You may also discontinue participation at any time without any penalty.

In addition, I am asking for permission to use your comments concerning these evaluations. These comments will be beneficial to understand parents' views. Your participation is also voluntary and confidential. Your refusal to participate will result in no penalty and you may discontinue participation at any time without penalty. You will not be identified by name.

If there are any questions concerning this research or the rights of the research subjects, you may contact the Graduate College, University of Northern Iowa, (319) 273-2748. All information will be kept for future reference. Thank you for your support.

Sincerely,
Beth Nordmeyer

I give my permission for my child's evaluations and for my comments about those evaluations to be used in this study.

Date_________________________ Name____________________________
Appendix B

Compiled Evaluations for lesson planning

Notes for needed instruction based on weekly evaluations for: _________

General Goal: To develop reading comprehension through the use of reading comprehension strategies.

Objectives: Jeremy B. Josh Katie Kenny

To Predict
Obsv. sheet

To Use Context
Running Record

To Use Prior Know.
Obsv. Sheet
Retelling

To Imagine
Obsv. sheet

To Vary Strategies
Running Record

To Retell Story Elements
Retelling
Compiled Evaluations
for lesson planning

Notes for needed instruction based on weekly evaluations for: 

General Goal: To develop reading comprehension through the use of reading comprehension strategies.

Objectives: 

<table>
<thead>
<tr>
<th></th>
<th>Aaron</th>
<th>Devin</th>
<th>Jeremy D.</th>
<th>Lance</th>
</tr>
</thead>
<tbody>
<tr>
<td>To Predict</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obsv. sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Use Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running Record</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Use Prior Know</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obsv. Sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retelling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Imagine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obsv. sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Vary Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running Record</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Retell Story Elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retelling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Compiled Evaluations
for lesson planning

Notes for needed instruction based on weekly evaluations for: 

General Goal: To develop reading comprehension through the use of reading comprehension strategies.

Objectives: Andy Bryce Casey Zeb Justin

To Predict
Obsv. sheet

To Use Context
Running Record

To Use Prior Know.
Obsv. Sheet
Retelling

To Imagine
Obsv. sheet

To Vary Strategies
Running Record

To Retell Story Elements
Retelling
Appendix C

OBSERVATION SHEET

Student ___________________________ Date __________________

General Remarks: ____________________________________________
_________________________________________________________________
_________________________________________________________________

Week's Summary

Behaviors being observed: 0 1-2 3 more

1. Predicts while reading:
   Tally: ______________________
   Comments: ______________________
   _______________________________________________________
   _______________________________________________________

2. Stimulates previous knowledge:
   Tally: ______________________
   Comments: ______________________
   _______________________________________________________
   _______________________________________________________

3. Imagining the story:
   Tally: ______________________
   Comments: ______________________
   _______________________________________________________
   _______________________________________________________
Appendix D

Self-Evaluation Sheet

Name: ___________________________ Date: ________________

Things I can do well:

__________________________________________________________________________________

(Teacher: Explain the following to the children.) Put your pencil on the first *. Read the sentence next to it. Draw an arrow to the no if you do not do what it says. Draw an arrow to the yes if you do what it says. The length of the line represents how strongly you answer yes or no.

1. I like to read. No * Yes
2. I feel good about the way I can read.
   No * Yes
3. I can tell you about what I have read.
   No * Yes
4. I use other words to help me find the meaning of a word.
   No * Yes
5. I can predict about what may happen in the story.
   No * Yes
6. I look at the illustrations for clues to the story.
   No * Yes
7. I imagine the story in my mind when I read.
   No * Yes
8. I read at home.
   No * Yes

Things I can improve:

__________________________________________________________________________________
Learning to Take a Running Record

Learning to take a running record can unsettle teachers. Those who are used to standardised tests and norms suspect the simplicity of the behaviour records, and so do people who do not like standardised testing.

There is not a lot to learn before you begin record taking, just a few conventions. There is no reason to study a new set of concepts or understand something new about the reading process. The first step is a matter of action. You set yourself the task of recording everything that a child says and does as he tries to read the book you have chosen. Once you begin such recording, and after about two hours of initial practice, no matter how much you might be missing, you have made a good start. The more you take the records the more you will notice about the child's behaviour. It is not a case of knowing everything first and then applying it. Try yourself out and you will begin to notice a few things that you have not noticed before. Practise some more and you will notice more. As your ear becomes tuned-in to reading behaviours and you gain control over the recording conventions your records will become more and more reliable.

I had been teaching reading and remedial reading for many years when I began my research on emergent reading behaviour. I am still humble about the fact that I had really never noticed self-correction behaviour until I started recording everything that children were doing. Then I found that I had been missing something that was very important.

What we are observing and recording is not something that is peculiar to the child who is learning to read. If I take some adult volunteers and ask them to read some ordinary everyday reading materials their reading behaviour can be broken down so that we can observe the same kinds of behaviour that you can observe in children's reading. A smudgy carbon or a bad stencil or a Churchill speech in i.t.a. or a misprint in the newspaper where the lines have been misplaced, such texts will break down the reading behaviours of these competent adults and one can observe self-correction, word-by-word reading and even the use of a pointing finger to locate themselves on the text. Everybody's reading behaviour can be broken down under difficulties.

Make a record of each child reading his three books or book selections. Use ticks for each correct response and record every error in full. A sample reading of 100 to 200 words from each text is required. This should take about 10 minutes. At the early reading level when the child is reading the very simplest texts the number of words may fall below 100 but if three texts are attempted (selected from caption books or first readers or teacher-made books)
or child-dictated text) this will be satisfactory.

A suggested format for a Running Record Summary Sheet can be found on page 109 and procedures for calculating accuracy and self-correction rates on page 115 of this book.

Some conventions used for recording

1. Mark every word read correctly with a tick (or check). A record of the first five pages of the Ready to Read book *Early in the Morning*, that was 100 percent correct, would look like this.

   Bill is asleep
   'Wake up, Bill,' said Peter.
   Sally is asleep.
   'Wake up, Sally,' said Mother.
   Father is shaving.

2. Record a wrong response with the text under it.

   Child: *home*
   Text: house

3. If a child tries several times to read a word, record all his trials.

   Child: *here* ____________ home
   Text: house

   Sounding out may be recorded in lower case, n-o-t and spelling the letters in capitals N-O-T.

4. If a child succeeds in correcting a previous error this is recorded as 'self-correction' (written SC).

   Child: *where* we ____________ when ____________ SC
   Text: were

   Example 3 does not end in self-correction.

5. If no response is given to a word it is recorded with a dash. Insertion of a word is recorded over a dash.

   No response: ____________
   Insertion: Child: *here*
   Text: house

6. If the child baulks, unable to proceed because he is aware he had made an error and cannot correct it, or because he cannot attempt the next word, he is told the word (written T).

   Child: *home* ____________
   Text: house  T

7. An appeal for help (A) from the child is turned back to the child for further effort before using T as in 6 above. Say 'You try it'.

   Child: _ A _ here
   Text: house _ _ T

8. Sometimes the child gets into a state of confusion and it is necessary to extricate him. The most detached method of doing this is to say 'Try that again', marking TTA on the record. This would not involve any teaching, but the teacher may indicate where the child should begin again.

   It is a good idea to put square brackets around the first set of muddled behaviour, enter the TTA, remember to count that as one error only (see page 20), and then begin a fresh record of the problem text. An example of this recording would be this.

   [ look said the headmaster ]

   Timothy went TTA to
   said R SC ____________
   went ____________

9. Repetition is not counted as error behaviour. Sometimes it is used to confirm a previous attempt. Often it results in self-correction. It is useful to record it as it often indicates how much sorting out the child is doing. 'R' standing for repetition, is used to indicate repetition of a word, with R, or R, indicating the number of repetitions. If the child goes back over a group of words, or returns to the beginning of the line or sentence in his repetition, the point to which he returns is shown by an arrow.

   Child: *Here is the home* R SC
   Text: ____________ Here is the house
Directional attack on the printed text is recorded by telling the child to ‘Read it with your finger’.

Left to right L → R
Right to left L → R
Snaking
Bottom to top B → T

For special purposes teachers or researchers may wish to develop their own conventions for scoring other behaviours which they notice. Some behaviours may be specific to, or important for, a particular teaching programme. For example, pausing can be recorded by a slash, / . Some researchers who have been concerned with the length of pausing have used a convention borrowed from linguistics which allows for pauses of four different lengths. These are quickly recorded as

/ / / /

I would caution against attention to pausing unless there is a special reason for wishing to record it. In research studies it has not yet yielded clear messages about the reading process. It adds little to the teacher’s interpretation of her record and may add confusion. It would be important not to read into a record of pausing interpretations for which there was no other evidence.

A running record from a child who is making many errors is harder to take and score but the rule is to record all behaviour, and analyse objectively what is recorded.

Reliability
Taped recordings of such reading observations taken from four children over the period of one year were available and were used to check on the reliability of such records (0.98 for error scoring and 0.68 for self-correction scoring, Clay 1966).

A number of trends became obvious during these reliability tests.

• For beginning readers, observers can take running records which give reliable accuracy scores with a small amount of training.
• The effect of poor observation is to reduce the number of errors recorded and increase the accuracy rate. As the observer’s skill in recording at speed increases, so the error scores tend to rise.
• To record all error behaviour in full, as against only tallying its occurrence, takes much more practice (but provides more evidence of the child’s strategies).
• Observations for poor readers require longer training to reach agreement on scoring standards because of the complex error behaviour.

Information is lost in the taped observation, especially motor behaviour and visual survey, but observation of vocal behaviour tends to be improved.

Reliability probably drops as reading accuracy level falls because there is more error behaviour to be recorded in the same time span.

For research work the most reliable records would be obtained by scoring an observation immediately following its manual recording, and re-checking immediately with a taped observation.

Analysing the reading record
From the running record of reading behaviour containing all the child’s behaviour on his current book, consider what is happening as the child reads.

Some conventions for scoring the records
In counting the number of errors, some arbitrary decisions must be made but the following have been found workable.

1 Credit the child with any correct or corrected words.
Child: to the shops Errors: 2
Text: for the bread Score: × × ×

2 There is no penalty for trials which are eventually correct.
Child: Want won’t went (SC) Errors: 0
Text: Went Self-correction: 1
Score: — — —

3 Insertions add errors so that a child can have more errors than there are words in a line.
Child: The train went toot, toot, toot Errors: 5
Text: The little engine sighed Score: × × × × × ×

4 However, the child cannot receive a minus score for a page. The lowest page score is 0.
The Early Detection of Reading Difficulties

5 Omissions. If a line or sentence is omitted each word is counted as an error.
   - If a page is omitted (perhaps because two pages were turned together) they are not counted as errors. Note that, in this case, the number of words on that page must be deducted from the Running Words Total before calculation.

6 Repeated errors. If the child makes an error (e.g. 'run' for 'ran') and then substitutes this word repeatedly, it counts as an error every time; but substitution of a proper name (e.g. 'Mary' for 'Molly') is counted only the first time.

7 Multiple errors and self-corrections. If a child makes two or more errors (e.g. reads a phrase wrongly) each word is an error. If he then corrects all these errors each corrected word is a self-correction.

8 Broken words. Where a word is pronounced as two words (e.g. a/way) even when this is backed up by pointing as if it were two words, this is regarded as an error of pronunciation not as a reading error unless what is said is matched to a different word. Such things as 'pitcher' for 'picture' and 'gonna' for 'going to' are counted as correct.

9 Inventions defeat the system. When the young child is creatively producing his own version of the story the scoring system finally breaks down and the judgement 'inventing' is recorded for that page, story or book.

10 'Try that again'. When the child is in a tangle this instruction, which does not involve teaching, can be given. It counts as one error and only the second attempt is scored.

11 Fewest errors. If there are alternate ways of scoring responses a general principle is to choose the method that gives the fewest possible errors as in B below.

A Child: We went for the bread
   Text: You went to the shop for the bread
   Score: × ✔ ✔ × × ✔ × ✔ ✔ × × × X
   Errors: 6

B Child: We went ___ for the bread
   Text: You went to the shop for the bread
   Score: × ✔ ✔ × × ✔ ✔ ✔ × × ✔ ✔ ✔ × × × X
   Errors: 4

Check directional movement

Ask the child to 'Read it with your finger.' While this may not be a desirable teaching instruction it is a necessary one for the observer to elicit evidence of directional movement. Record which hand was used, on which page, and the direction of movement.

In a study of children's early learning of directional movement across English texts a common progression was noted (Clay, 1982). There was an early period of confusion as the children tried to orient to the spatial characteristics of the open book. Then there was a period when the child seemed to prefer to use a particular hand for pointing to any text. Finally a more flexible set of behaviour emerged when the child could use either hand on either page without having to pay much attention to direction. As these stages were worked through, sometimes rapidly and sometimes over several months, lapses from directional behaviours were observed. Children might go from right to left or even from bottom to top. Left and right handed children showed similar kinds of behaviours.

Three groups of children have difficulty as beginning readers in disciplining their behaviour within the directional constraints of written language.

- The first group are children who have poor motor coordination or who are inattentive to where their bodies are and how they are arranging their movements.
- The second group are fast-reacting, impulsive children who act before they think and who have great difficulty in governing their responses within any constraints. They can very readily settle into undesirable patterns of responding.
- A third group of beginners at risk are those who do not like to try because they might make a mistake. As the development of directional behaviour involves exploring two-dimensional space, being wrong, and discovering how to behave correctly, children who are too tense, inhibited or timid, may be reluctant to try out a range of directional behaviour. They take longer to learn to discard the poor responses and retain the good ones.

The technique of asking a child to 'Read it with your finger' will only reveal directional behaviour on the gross schema of line scanning. Beyond this there must be some very important visual perception learning to be done. It relates to the scanning of letters and clusters of letters. There are further important orientation behaviours to be learned to do with what the eyes are attending to, in what order, which will not be picked up in observations of pointing behaviour.

Record your observations and comments on directional movement on your summary sheet (pp.109, 111). Any lapse from appropriate directional behaviour is important and should be noted. We are not merely concerned with
the child who habitually moves in the wrong direction, but rather with the child who is inconsistent, or in the process of learning directional control shown by lapses from time to time.

Calculate the error rate
Compare the number of errors with the number of running words. Does the child read his book with one error in every five running words of text (which is poor) or is it more like one in twenty (which is good)? Record on the Summary of Running Record Sheet.

Calculate the percentage of errors (see Conversion Table page 115). If there is more than 10 percent of error in the record rate this is a 'hard' text for this child. (For the average child there is movement from 90 percent accuracy when he is first promoted to a book to 95 percent or more as he completes his learning on that book.)

When the child reads a book with less than 90 percent accuracy it is difficult for him to judge for himself whether his attempt at a word is a good one or a poor one. He needs easier material which he can attempt at a rate of not more than one error in ten words at the time he begins the new book. The reading text should use language that comes readily to him. In the very earliest stages it is sometimes necessary to repeat the text until he has almost memorised it, but not quite. Then it will come readily to the tip of his tongue. It is as if the words he needs are stored in the depths of his memory and have to be assisted to float to the surface. The child's own dictated stories provide good reading texts for just this reason — the words and construction of the text should be readily recalled.

If the text is in a different style from that which the child usually reads his error rate will increase because he is predicting from the baseline of old expectations which are inappropriate for the present text.

Error behaviour
To read a continuous text the child must use a variety of skills held in delicate balance. Specific weaknesses or strengths can upset that balance. There are some questions about the errors for a particular child that can guide the teacher's analysis of the behaviour record. (See also Clay, 1979.)

Oral language skills. Are these good enough to make the reading of this text possible? (For instance, could the child repeat the sentences of the text if you asked him to, one by one?) Or, is his language so fluent that the coordination of visual perception and motor movement with language is difficult?

Speed of responding. The rate at which a child reads and the time spent on pausing and processing cues are at this level poor indicators of the child's progress. One child may read with the fluency of oral language but may be a poorer reader than another child who pauses and engages in much self-correction behaviour. At this particular stage in reading progress it is good for the child making average progress to be concerned about error and to rectify error if possible. It is poor to maintain fluency and not to notice that one has made an error.

Fast responding can be an indication that language is dominating the reading process allowing for little visual search to take place.

What cues does he depend on?
- Does the child use meaning? If what he reads makes sense, even though it is inaccurate, then he is probably applying his oral language knowledge to his reading.
- Is what he says grammatical? If it is, his oral language is influencing his responding. If it is not, there may be two reasons. Perhaps his language skill is limited and his personal 'grammar' does not contain the structures used in his reading book. Or, if he is paying close attention to detail, or to word by word reading, he may not be allowing his control over English syntax to influence his choices.
- Does he use visual cues from the letters and words?
- Does he read word by word as if recalling each word from a memory bank, unrelated to what has gone before? He may not realise that reading is like speaking, and that his language behaviour is a rich source of help in choosing correct reading responses.

To work out whether the child is responding to the different kinds of cues that could be used you need to look at every error that the child makes and ask yourself 'Now what made him say that?' 'Did he miss out on the visual cues?' 'Was he ignoring meaning?' It is misleading if you do this selectively; you must analyse every error and count those that show this or that kind of cue. You want to be able to conclude, on sound evidence, that 'He pays more attention to visual cues than to meaning', or 'He is guided by structure and meaning but does not search for visual cues'. It is only when you go to the trouble of analysing all the errors that you really get any indication of what his strategies are on reading.

When teachers are familiar with taking running records they may want to write M for meaning, S for structure and V for visual cues on the record form and to record, by circling, which cues the child was using. (See record sheets p. 109). Notice that what you are recording in this case is your best guess: you cannot know what cues the child used. A record may show one, two or three types of cues used on any one error. If you write M S V alongside each error or self-correction and circle the cues you think the child used, the uncircled letters will then show the cues neglected.
Consider the error first. What cues up to that error was the child using? Think only of the information the child had before the error occurred. Then consider the self-correction. What extra information did the child use in the self-correction?

Enter comments on the Analysis of Errors. (Summary of Running Record Sheet.)

Cross-checking strategies
Can the child check one kind of information with another? Can he get movement and language occurring together in a linked or coordinated way? Does he check on language prediction by looking at some letters? Can he hear the sounds in a word and check whether the expected letters are there? A child with outstanding memory for what he hears or with very fast language production often has difficulty in slowing up enough to enable him to learn the visual discriminations. Yet good readers search for cues from different sources which confirm a response. (See pages 72 to 74 and Clay 1979 for further discussion of these reading behaviours.)

Self-correction
Observe and enter in the running record any self-correction behaviour. The child discovers cues that tell him something is wrong. He is aware that a particular message is to be communicated and tries to discover this by using cues. Efficient self-correction behaviour is an important skill in good reading. Calculate the self-correction rate (see page 19). A self-correction rate of one in three to five errors is good but one in twenty errors is a very low rate. However the prognosis is good, because self-correction does exist!

Self-correction rates vary greatly. This is because they are not absolute scores: they are always relative measures. They vary with text difficulty, with error rate, with accuracy, and with effort. They cannot be understood unless they are interpreted together with text difficulty and accuracy scores.

If self-correction is evident but inefficient it is a good prognosis. Its absence in a record which contains errors is a danger sign. A child who is making errors and is not aware of this, or who makes no attempt to correct himself, is in difficulties. He is not aware of the need to decode a precise message or he is not aware of the existence of cues, or he does not know how to use them, or he does not try to solve the problem.

If a child engages in a confusion of unsuccessful attempts to solve his errors he needs to learn better bases for making his decisions. His teacher must deliberately teach some priorities like 'Sound the first letter', 'Go back to the beginning of the line', 'What would make sense?' — whichever she judges to be the technique with the highest pay-off in terms of progress, for this child at this time. (See page 73.)

Linda was 5:9 when she was reading the book which gave rise to the example of reading behaviours on the next page. (It was not a very helpful book for her level of reading.) You might think that she was a poor reader. Yet when you think about what is going on in this record, and how many things she is trying to do, and what kinds of cues she is testing out, you can see that she really is working hard to relate one kind of information to another. This is a very interesting record of her behaviour, showing how active she is in searching, and checking. In time she must become more efficient at doing these things.

Appendix F

Running Record for ___________________________ Date ____________

Book ________________________________ Author ________________
Appendix G

Story Retelling:
Overall Organization Rating Sheet

RATE 3: Events that accurately represent the author’s events
Events that are told according to the author’s sequence
A beginning that matches the author’s
A well-defined story problem that matches the author’s
A conclusion that wraps up the story according to the original
No irrelevant details

RATE 2: Some story events that change the author’s events
Some events that are told out of order
A beginning and conclusion that roughly match the author’s
Action that rambles without clearly defining the author’s
central problem
Few irrelevant details

RATE 1: Many events that do not appear in the author’s story
Few events recalled
A beginning and/or ending that differs from the author’s
Omission of the author’s central story problem
Many irrelevant events

RATE 0: No main idea or point to the story
Practically no events recalled
An overall impression of disorder because of jumbled
arrangement of ideas
No beginning, middle or end
Many irrelevant details

Appendix H

Running Record and Retelling Accumulative Sheet

<table>
<thead>
<tr>
<th>Title</th>
<th>Date</th>
<th>ident corr</th>
<th>context match</th>
<th>Retelling</th>
<th>Comments</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>% words</th>
<th>% correct plus</th>
</tr>
</thead>
</table>


Appendix I

END OF YEAR ACCUMULATIVE EVALUATION SHEET

Name: ____________________________

Week of:  

<table>
<thead>
<tr>
<th>Date</th>
<th>Self-Evaluation</th>
<th>Running record</th>
<th>Retelling</th>
<th>Observ.</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix J

Evaluation Management
Teacher should note whether each of the following evaluations were completed. A comment should be made if it was not completed.

<table>
<thead>
<tr>
<th>Form of Evaluation:</th>
<th>Student Journal and Retellings</th>
<th>Running Records*</th>
<th>Observ.*</th>
<th>Self Eval.**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Running records/retellings and observations are done on alternating weeks.  
**Self-evaluations are done monthly.
Appendix K

PARENT'S QUESTIONNAIRE

Student's Name: ______________________ Date: ____________

Parent's Name: ______________________

Please comment on the following evaluations which were used in your child's Chapter 1 Reading room. First choose the letter which best represents your opinion about each type of evaluation and then write any other comments you would like to share.

Use the following corresponding letters to the responses in rating the following evaluations.

(A) I thought this information was very beneficial.
(B) I thought this information was helpful.
(C) This information did not matter to me.
(D) This information was not beneficial to me.
(E) I did not understand this information.

1. The student's self-evaluations. A B C D E
   Comments:

2. The teacher's observations regarding your child's reading. A B C D E
   Comments:

3. The information in your child's portfolio. A B C D E
   Comments:

4. The information in your child's journal. A B C D E
   Comments:
STUDENT'S INTERVIEW

Student's Name: ___________________________ Date: __________________

1. Do you know what is in your portfolio?

2. If so, describe it to me, telling me what it tells you.
TEACHER'S QUESTIONNAIRE Part I

Teacher's Name: __________________________ Date: ________________

Please comment on the following evaluations which were used in your students' Chapter 1 Reading room. First choose the letter which best represents your opinion about each type of evaluation and then write any other comments you would like to share.

Use the following corresponding letters to the responses in rating the following evaluations.

(A) I thought this information was very beneficial.
(B) I thought this information was helpful.
(C) This information did not matter to me.
(D) This information was not beneficial to me.
(E) I did not understand this information.

1. The students' self-evaluations.
   Comments: A B C D E

2. The Chapter 1 teacher's observations regarding your students' reading.
   Comments: A B C D E

3. The information in the students' portfolios.
   Comments: A B C D E

4. The information in your students' journals.
   Comments: A B C D E
TEACHER QUESTIONNAIRE PART II

Teacher's Name: ____________________ Date: ________________

Please comment on how you used the information from the following types of evaluations used in your students' Chapter 1 classroom. First choose the number which best represents your use of the evaluation and then explain why that evaluation was useful or why it was not useful.

Use the following corresponding numbers to the responses in rating how you used the evaluations.

(A) This evaluation was extremely useful.
(B) This evaluation was quite useful.
(C) This evaluation was moderately useful.
(D) This evaluation was of little use.
(E) This evaluation was of no use.

1. The students' self-evaluations. A B C D E
   Comments:

2. The Chapter 1 teacher's observations regarding your students' reading. A B C D E
   Comments:

3. The information in the students' portfolios. A B C D E
   Comments:

4. The information in your students' journals. A B C D E
   Comments:
# Appendix L

## END OF YEAR ACCUMULATIVE EVALUATION SHEET

**Name:** Casey

<table>
<thead>
<tr>
<th>Week of</th>
<th>Self-Evaluation</th>
<th>Running record</th>
<th>Retelling</th>
<th>Observ.</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/28</td>
<td>remember - no!</td>
<td>use words - no!</td>
<td></td>
<td></td>
<td>1/28 I predict 1 imagine stories</td>
</tr>
<tr>
<td></td>
<td>use words in context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2/4</td>
<td></td>
<td>64% / 67%</td>
<td></td>
<td></td>
<td>2/11 2 predict I no how to read</td>
</tr>
<tr>
<td>2/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2/11 0 imagine to read</td>
</tr>
<tr>
<td>2/18</td>
<td></td>
<td>84% / 86%</td>
<td></td>
<td></td>
<td>2/18 1 know.</td>
</tr>
<tr>
<td>2/25</td>
<td>remember - no</td>
<td>use words - yes</td>
<td></td>
<td></td>
<td>2/25 daily happenings</td>
</tr>
<tr>
<td>3/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3/11 0 write nothing</td>
</tr>
<tr>
<td>3/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3/11 1 predict refused</td>
</tr>
<tr>
<td>3/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3/18 0 related</td>
</tr>
<tr>
<td>3/25</td>
<td>remember - yes</td>
<td></td>
<td></td>
<td></td>
<td>3/25 0 prior know. daily events</td>
</tr>
<tr>
<td>4/8</td>
<td></td>
<td>99% / 100%</td>
<td></td>
<td></td>
<td>4/15 3 predict 0 imagine 1 prior know.</td>
</tr>
<tr>
<td>4/15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4/15 2 imagine daily events</td>
</tr>
<tr>
<td>4/22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4/22 1 prior know.</td>
</tr>
<tr>
<td>4/29</td>
<td>remember - yes</td>
<td></td>
<td></td>
<td></td>
<td>4/29 3 predict 3 imagine</td>
</tr>
<tr>
<td>5/6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5/6 no time</td>
</tr>
<tr>
<td>5/13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5/13 no time</td>
</tr>
</tbody>
</table>

*Full good - no every time all others - all usea*
## Appendix M

### Running Record for Casey

<table>
<thead>
<tr>
<th>Book</th>
<th>Author</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mike's New Bike</strong></td>
<td><strong>Graydanua</strong></td>
<td><strong>4-9-91</strong></td>
</tr>
<tr>
<td>V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>Bike goes</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>But get</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>T V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td><em>like am</em></td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V and</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>are in</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>He</td>
<td>V V V V V V</td>
<td>V V V V V V</td>
</tr>
<tr>
<td>His</td>
<td>sc</td>
<td>V V V V V V</td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td><em>looks like this</em></td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
<tr>
<td>V V V V V V</td>
<td>V V V V V V</td>
<td></td>
</tr>
</tbody>
</table>