INFORMATION TO USERS

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

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Profile analysis of error types and strategy use on a standardized, reading comprehension test: Average readers versus poor readers

Johnston-Anderson, Karen Marie, Ph.D.
The Ohio State University, 1992

Copyright ©1992 by Johnston-Anderson, Karen Marie. All rights reserved.
PROFILE ANALYSIS OF ERROR TYPES AND STRATEGY USE ON A STANDARDIZED, READING COMPREHENSION TEST: AVERAGE READERS VERSUS POOR READERS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Karen Marie Johnston Anderson, B.S., M.A., N.C.S.P.

*****

The Ohio State University

1992

Dissertation committee: Judy L. Genshaft, Ph.D. Sandra McCormick, Ph.D. Jack A. Naglieri, Ph.D.

Approved by

Adviser

College of Education
To my parents John and Jewel Johnston,

and my brother John E. Johnston, for their

love and support
ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Dr. Sandra McCormick for her unrelenting assistance and time commitment throughout the course of this project. In addition, I would like to thank Dr. McCormick and Dr. Judy Genshaft for providing me with several exceptional educational and employment opportunities throughout my graduate career. The author would also like to acknowledge the contributions of Dr. Judy Genshaft and Dr. Jack Naglieri in the development of my knowledge and skills in the area of school psychology. Thanks also go to my friend and colleague, Jim Broyles, for his assistance in reliability assessment of the data in this study.
VITA

October 3, 1962..............Born - Portland, Oregon

1984.........................B.S. Carroll College, Waukesha, Wisconsin

1985-1986.....................Graduate Research Associate, Ohio State University. Research in Reading Disabilities, Buckeye Youth Center, Columbus, Ohio

1986-1987.....................Graduate Research Associate, Ohio State University. Reading Clinic Supervisor

1987-1988.....................Intern School Psychologist, South-Western City Schools, Grove City, Ohio

1988-1990.....................Graduate Research Associate, Ohio State University. School Psychology Clinic Supervisor

1988-present..................Psychology Assistant, Private Practice, Dr. Jolie Brams and Associates, Columbus, Ohio

FIELD OF STUDY

Major Field: Education

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I.   INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Prevalence of Standardized Testing</td>
<td>1</td>
</tr>
<tr>
<td>Issues Regarding Standardized Comprehension Assessment</td>
<td>3</td>
</tr>
<tr>
<td>The At-Risk Population</td>
<td>9</td>
</tr>
<tr>
<td>II.  REVIEW OF THE LITERATURE</td>
<td></td>
</tr>
<tr>
<td>Issues Related to Standardized Testing</td>
<td>13</td>
</tr>
<tr>
<td>Relationship Between Teacher Ratings and Test Scores</td>
<td>14</td>
</tr>
<tr>
<td>Test Content and Curricular Mismatch</td>
<td>21</td>
</tr>
<tr>
<td>Group-Administered Comprehension Tests and Students At Risk</td>
<td>22</td>
</tr>
<tr>
<td>Conclusion</td>
<td>24</td>
</tr>
<tr>
<td>Theories on the Nature of the Reading Process</td>
<td>25</td>
</tr>
<tr>
<td>LaBerge and Samuels Model</td>
<td>26</td>
</tr>
<tr>
<td>Goodman Psycholinguistic Model</td>
<td>26</td>
</tr>
<tr>
<td>Interactive Models</td>
<td>27</td>
</tr>
<tr>
<td>Rumelhart's Model</td>
<td>28</td>
</tr>
</tbody>
</table>

v
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanovich's Model</td>
<td>29</td>
</tr>
<tr>
<td>Just and Carpenter's Model</td>
<td>30</td>
</tr>
<tr>
<td>Other Models</td>
<td>31</td>
</tr>
<tr>
<td>Metacognition and Reading</td>
<td>33</td>
</tr>
<tr>
<td>Factors Influencing Student</td>
<td></td>
</tr>
<tr>
<td>Performance on Standardized</td>
<td></td>
</tr>
<tr>
<td>Comprehension Tests</td>
<td>38</td>
</tr>
<tr>
<td>Specific Reading Problems</td>
<td>38</td>
</tr>
<tr>
<td>Word Recognition and word</td>
<td></td>
</tr>
<tr>
<td>identification</td>
<td>39</td>
</tr>
<tr>
<td>Meaning Vocabulary</td>
<td>41</td>
</tr>
<tr>
<td>Comprehension</td>
<td>42</td>
</tr>
<tr>
<td>Strategic Factors</td>
<td>47</td>
</tr>
<tr>
<td>Reading Strategies</td>
<td>47</td>
</tr>
<tr>
<td>Test-Taking Strategies</td>
<td>54</td>
</tr>
<tr>
<td>Mode of Test-Taking</td>
<td>60</td>
</tr>
<tr>
<td>Mode of Reading</td>
<td>61</td>
</tr>
<tr>
<td>Group versus Individual Administration</td>
<td>62</td>
</tr>
<tr>
<td>Research-Related Methodological Issues</td>
<td>65</td>
</tr>
<tr>
<td>Summary of the Literature Review</td>
<td>71</td>
</tr>
<tr>
<td>Research Questions</td>
<td>74</td>
</tr>
<tr>
<td>Errors Between Groups: Poor</td>
<td></td>
</tr>
<tr>
<td>Readers versus Average Readers</td>
<td>74</td>
</tr>
<tr>
<td>Strategy Use Between Groups: Poor</td>
<td></td>
</tr>
<tr>
<td>Readers versus Average Readers</td>
<td>75</td>
</tr>
<tr>
<td>Answer Changes During Individual Test</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>77</td>
</tr>
<tr>
<td>III. METHOD</td>
<td>79</td>
</tr>
<tr>
<td>Subjects</td>
<td>79</td>
</tr>
<tr>
<td>Setting</td>
<td>85</td>
</tr>
<tr>
<td>Experimenter</td>
<td>87</td>
</tr>
<tr>
<td>Instruments and Materials</td>
<td>87</td>
</tr>
<tr>
<td>Intelligence Test</td>
<td>87</td>
</tr>
<tr>
<td>Matrix Analogies Test</td>
<td></td>
</tr>
<tr>
<td>Expanded Form</td>
<td>87</td>
</tr>
<tr>
<td>Reading Test</td>
<td>89</td>
</tr>
</tbody>
</table>
Limitations of this Study.....................156
Directions for Future Research...............160
Summary..................................163

VI. APPENDICES..........................169
    A. Letter to Parents and
       Consent Form.......................170
    B. Individual Observation Sheet
       and Strategy Checklist...........174
    C. Error Coding Sheet................176
    D. Checklist for Reliability of
       Experimenter’s Procedures......178

LIST OF REFERENCES........................179
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Frequencies, Means, Standard Deviations, and Percentages of Word Identification (WI), Meaning Vocabulary (MV), and Comprehension (COMP) Errors by Group</td>
<td>119</td>
</tr>
<tr>
<td>2. A Description of Strategies</td>
<td>125</td>
</tr>
<tr>
<td>3. Total Frequency and Percentage of Strategy Use by Group</td>
<td>126</td>
</tr>
<tr>
<td>4. Strategy Use When Students Answered Questions Correctly</td>
<td>130</td>
</tr>
<tr>
<td>5. Strategy Use When Students Answered Questions Incorrectly</td>
<td>135</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A comparison of Profiles (Mean Transformed Proportions) of Error Types by Group</td>
<td>122</td>
</tr>
<tr>
<td>2. Mean (Transformed Proportions) Scores of Average and Poor Readers by Error Type</td>
<td>123</td>
</tr>
<tr>
<td>3. A Comparison of Profiles (Mean Transformed Proportions) of Strategy Use by Group</td>
<td>128</td>
</tr>
<tr>
<td>4. Mean Transformed Proportions of Average and Poor Readers for Overall Strategy Use</td>
<td>129</td>
</tr>
<tr>
<td>5. A comparison of Profiles (Mean Transformed Proportions of Strategy Use by Group for Correct Answer Selection</td>
<td>132</td>
</tr>
<tr>
<td>6. Mean Scores of Average and Poor Readers for Strategy Use when Students Selected Correct Answers</td>
<td>133</td>
</tr>
<tr>
<td>7. A Comparison of Profiles (Transformed Mean Proportions of Strategy Use by Group for Incorrect Answers</td>
<td>136</td>
</tr>
<tr>
<td>8. Mean Transformed Proportions for Average and Poor Readers for Incorrect Answer Responding</td>
<td>138</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

The purpose of this study was to investigate and categorize the types of errors and strategies educationally at-risk students employ when taking a standardized reading comprehension test as compared to students not classified as at-risk. Through a combination of individual protocol analysis and analysis of retrospective reports, a student's test-taking difficulties, reading strategies, and misconceptions can be shown with regard to test performance. Profiles of error types and strategy use were then compared between the two groups in this study. This chapter introduces the reader to issues concerning standardized assessment of reading comprehension, the at-risk student population, and the importance of research in these areas.

Prevalence of Standardized Testing

The administration of standardized group achievement tests in the schools of the United States has reached an all time high. Over the past 10 years there has been a
significant increase in the amount of testing; it has been estimated that testing has increased at the rate of 10% to 20% each year (Haney & Madaus, 1989). In a survey by Neill and Medina (1989) it was found that public schools in the United States administered 105 million standardized tests to 39.8 million students during the 1986-1987 school year, which is more than 2.5 standardized tests per student. Neill and Medina also noted that standardized testing is most prevalent in the southern states and in large urban school systems--locales that contain large numbers of minority students and low income families.

Valencia and Pearson (1987) attribute the increase of standardized test usage to the movement toward teacher and school accountability. They also cite several national reports on the educational status of the United States expressing concern with the reading status of the nation (Education Commission of the States, 1983; Valencia & Pearson, 1987). In his State of the Union Address, January 31, 1990, President George Bush emphasized the importance of raising the achievement of the nation's school children. President Bush stated, "We must assess our children as an indicator of academic knowledge and as a gauge for the success of our nation's schools." This statement by the President of the United States suggests
that standardized testing will continue to play an important role in evaluating the effectiveness of our entire educational system.

The present study is specifically concerned with the measurement of reading comprehension by standardized, group-administered achievement tests. Group-administered reading comprehension tests are multiple-choice in nature, and according to Gambrell (1985), 70% of the tests now being used to measure reading competence in the schools are multiple-choice tests. According to Johnston (1984), the main reasons for frequent use of group tests are the ease of administration and scoring, and the ability to compare children across groups.

**Issues Regarding Standardized Comprehension Assessment**

Despite the prevalence of standardized test use, and no indication of their use diminishing, there are many difficulties when trying to objectively measure reading comprehension. According to Farr (1969), Farr and Carey (1986) and Johnston (1983), group-administered, standardized reading tests provide only a limited measure of a student's comprehension ability. Choosing the correct answer to a question is only one indicator of comprehension, and a major disadvantage of a multiple
choice format is the student’s inability to display other comprehending behaviors, such as justifying an answer (Readence & Moore, 1983). Furthermore, no information is given that distinguishes a student’s problem with word recognition versus use of erroneous comprehension strategies. Other criticisms of standardized reading tests include the lack of similarity of the short passages to real reading, and that passage content and style favor the cultural majority (Johnston, 1984).

Proponents of standardized tests often disagree with these claims, by noting the high correlation between achievement and cognitive ability measures (Stenner, Jackson, Horaban, Smith & Smith, 1988). (It should be noted, though, that the latter could also be construed as a weakness, for it could be argued since the tests are highly correlated, they lack specificity in measuring a certain skill). Researchers have also found a high degree of correlation between teacher judgment and standardized test scores (Hoge & Coladarci, 1989; Kellaghan, Madaus, & Airasian, 1982; Pedulla, Airasian, & Madaus, 1980), as well as a correlation of the scores with school grades (Hopkins & Stanley, 1981). Validity of standardized tests has also been demonstrated with culturally different children. It was found in Hochman’s (1973) research that
no significant difference existed between Black and Caucasian children when administered two standardized tests: one in Black vernacular dialect and one in Standard English.

Although the pros and cons of standardized tests are debated, one criticism of standardized comprehension tests seems to go unchallenged. Standardized comprehension tests have failed to incorporate new theoretical advances that focus on reading as an interactive, strategic process (Aaronson & Farr, 1988; Farr & Carey, 1986; Valencia & Pearson, 1987; Wood, 1988). Theoretical advances in reading have moved from bottom-up processing models (e.g. LaBerge & Samuels, 1974) and top-down processing models (e.g. Goodman, 1970) to an interactive stance (e.g. Rumelhart, 1976; Stanovich, 1980; 1991). In brief, bottom-up processing begins with the printed information and moves hierarchically into the brain resulting in comprehension. Top-down processing begins with the reader's intentions and hypotheses testing where the reader attends to printed matter only to the extent to absolve his remaining uncertainty. In contrast, interactive theories of reading processing posit that readers simultaneously attend to both cognitive and sensory stimuli which results in simultaneous decoding and
comprehension. These theories will be discussed further in the following chapter.

Testing theories of reading processing through analysis of the ongoing reading act can add empirical results that either supports or refutes these theories. However, measuring the reading process itself is difficult and the methodology is controversial. Reading comprehension tests have often been criticized for failure to measure the reading process, since many tests use a multiple-choice format. The problem of failing to measure reading processes is not new. Rusch and Stoddard (1927) noted that reading a paragraph and finding the answer to a question is a matching task that ignores the reasoning process. Johnston (1984) asserted that when a person simply answers multiple-choice questions after reading, there is no information on the reasoning process, hypotheses testing, or misconceptions underlying the individual's test performance.

Farr, Pritchard, and Smitten (1990) also question whether multiple-choice questions actually assess a reader's understanding of what has been read. They assert that multiple-choice reading tests may reinforce the notion of an inherently correct meaning the reader must determine from the text. Farr et al. criticize multiple
choice comprehension tests because it may be possible to respond correctly to an item without fully reading or comprehending the accompanying reading selection. They contend that "test-taking strategies are directed by the reader's goal to get an answer that is acceptable rather than to understand the reading selection" (p. 210). Although standardized, multiple-choice comprehension tests have been criticized as failing to measure reading processes, they continue to be widely used.

Farr et al. (1990) stated that there is little information about what an examinee actually does while taking a multiple-choice comprehension test, and it is possible to go beyond the traditional question and answer method of assessment. According to Johnston (1984) during an examinee's oral reading, useful knowledge can be gained through analysis of an individual's reading errors. Patterns of errors and other behaviors can clarify misconceptions that produce errors in comprehension. In regard to multiple-choice tests, Johnston (1983) stated,

If selection of a specific alternative by a reader indicates a particular reading strategy or problem, then we can begin to look at patterns of similar responses, and the assessment can become more efficient. This legitimizes the questions, "did the reader get it right for the right reason"? and "why did he get it wrong"? because one could examine patterns of responses to correct and
incorrect alternatives and make inferences about the strategies being used. (p. 59)

This strategic approach would detect students' misconceptions and offer insights into their performance (Johnston, 1984). Johnston suggests that with an individual analysis of mistakes, error patterns can be detected, and if they can be specified, then diagnostic progress can be made in the field of testing. Therefore, if an individual's performance on a standardized test was subjected to careful analysis of why certain answers were chosen, then the strategies, misconceptions, and other reasons for erroneous answer choices could be categorized and analyzed, and reasons for a student's low performance can be obtained. This method would result in a more process-oriented assessment.

Similarly, Farr et al. (1990) indicated a need for research focusing on a description of the process readers use during multiple-choice testing. This information would shed light on the validity of multiple-choice tests in reading comprehension assessment, as well as provide valuable reading process information.

Baker and Brown (1984) discussed the increasing interest of developmental and cognitive psychologists in a child's metacognitive status, that is, the knowledge and
control a child has over cognitive activities, such as reading. Metacognition is defined by Baker and Brown as "the self-regulatory mechanisms used by an active learner during an ongoing attempt to solve problems" (p. 361). They also note that metacognition can be thought of as "development and use of compensatory strategies" (p. 364) to solve a problem, which have obvious implications for the study of reading comprehension. Baker and Brown assert that "few studies have examined the metacognitive experiences of their subjects during reading" (p. 365), therefore it is an area in which research is important. In a recent review of articles related to metacognition and reading, Paris, Wasik, and Van der Westhuizen (1988) stated, "the database of metacognition and reading research rests on relatively few empirical studies published in refereed journals. Given the importance of this topic, we hope the database continues to expand" (p. 145).

The At-Risk Population

Since standardized, group-administered tests are widely used in inner city public schools (Neill & Medina, 1989), an individual reading comprehension analysis that
detects the strengths and weaknesses of a sample from this population can add insight into the problems of students classified as "at-risk." Although there are various definitions for the term at-risk, Slavin and Madden (1989) define an individual "at-risk" as one who is in danger of failing to complete his or her education and who lacks an adequate level of basic skills. In addition, they list risk factors such as retention in a grade, behavior problems, poor attendance, low socioeconomic status, and attendance at a school with a large number of poor students. According to Smith-Burke (1989) and MDC, Inc. (1988), although the minority and poor make up the majority of the at-risk population, the single common characteristic of at-risk youth is not race or economic disadvantage, but low scores on tests of basic skills, such as reading, writing, and math. Therefore, the importance of targeting youth that are low in socioeconomic status and low achieving would aid in understanding of the at-risk population.

Despite the importance of research using an at-risk population, most studies about cognitive processes in reading have been conducted with white, middle class children or college populations (Smith-Burke, 1989). Smith-Burke asserts that there is relatively little data
targeted specifically at literacy behaviors of linguistically and culturally different populations, particularly for low socioeconomic status minority groups. This is unfortunate since many would agree with Cuban (1989) who stated that the future of urban schools is the primary issue facing the educational system today, and many reforms seemed to have bypassed urban schools.

Given the prevalence and controversy of group-administered, multiple choice, standardized comprehension tests, and the paucity of data-based literature on at-risk students and reading processing, this study was proposed. Information is needed to further understand the strengths and weaknesses of at-risk children when taking standardized, group-administered, reading comprehension tests. The importance of obtaining a clearer picture of why students are performing poorly is necessary for growth in the field of assessment and reading comprehension instruction. Changing a test that measures only a product of reading and transforming it into a test that measures both product and process would be more consistent with theoretical advances in the area of reading. Examination of a student's reading and test-taking procedures using a combination of verbal report and protocol analysis methods, may shed light on which theories explain reading
behavior and reading problems more accurately. In addition, information from this study may offer implications for comprehension assessment and instruction.

Several questions can be raised when investigating assessment of reading errors and strategies. For instance, what error types and strategies occur most frequently on standardized reading comprehension tests? Is there a difference between groups in the resulting profiles of error types and strategy use? If so, are some error types and strategies more detrimental or more effective? These questions were addressed in this study, and are stated explicitly in the form of research questions in the following chapter.
CHAPTER II

REVIEW OF THE LITERATURE

The literature review is divided into five main sections. The first section includes research on standardized, group-administered tests, and is titled Issues Related to Standardized Testing. Special emphasis is given to tests of reading, and the validity issue of reading tests. The second section, titled Theories on the Nature of the Reading Process outlines three major theoretical models of reading. Section three titled Metacognition and Reading describes developments in research on knowledge of reading strategy use and regulation of cognition. The fourth section titled Factors Influencing Student Performance on Standardized Tests, addresses specific reading problems, strategic factors, and how mode of reading and test administration may impact a student's performance on a comprehension test. In section five, Research-Related Methodological Issues, advantages and disadvantages of the use of verbal report data are discussed in relation to reading research.
Issues Related to Standardized Testing

Ebel (1976) described the paradox of educational testing. He stated, "While assessments of achievement and competence are being more urgently called for and more widely employed than ever before, tests are at the same time, being more sharply criticized and more strongly opposed" (p. 1). Ebel's statements continue to be echoed today (Farr, Pritchard and Smitten, 1990). Hopkins, George and Williams (1985) also noted the paradox and stated, "Standardized achievement tests have been characterized as guessing games on the one hand and as definitive standards for educational excellence and accountability on the other" (p. 177). According to Kellaghan, Madaus and Airasian (1982), much of the criticism about testing seems based on political and ideological positions, and much of it emanates from people outside the testing profession. Furthermore, these debates are often held in the public arena, rather than within the boundaries of scholarly publications. Although the validity of standardized, multiple-choice tests have been criticized on several grounds, empirical investigations of teacher judgment ratings and test scores have largely supported the concurrent and criterion-related validity of the tests. In addition, concerns
regarding mismatch between test content and the school curriculum may be unwarranted.

According to Farr and Carey (1986) and Anastasi (1988) tests do not possess validity as some inherent characteristic. Validity depends on the purpose for which the test is constructed and its intended use. Farr and Carey stressed that reading test performance is only a sample of reading behavior, and what actually constitutes reading is a controversial issue. For example Farr and Carey stated,

Validity is the most important characteristic of a test, but the search for validity evidence is much more elusive than this definition suggests. Indeed, the search for validity is a search for the true behavior being measured. Since there is still much to be learned about reading, it is reasonable to accept the notion that a totally valid reading test does not exist. Considered from this perspective, it is best to think of a test in terms of degrees of validity. A test may measure some aspect of reading and, along with other observations of performance, may provide some insight into a person’s reading ability. (p. 137)

Anastasi (1988) emphasized that group-administered achievement batteries may differ in technical level of test development, but as a group they must conform to high standards of test development. Most major achievement batteries have large, representative norm samples, and provide information on reliability and content validity.
She also noted that test development consists of several procedures to assure that item content is appropriate, and that sex and ethnic bias are avoided.

Relationship Between Teacher Ratings and Test Scores

A number of empirical studies have compared teacher ratings and standardized test scores. Hopkins, George and Williams (1985) noted that if there is a correlation between teacher judgment and test scores, it provides evidence of test validity. According to Hoge and Coladarci (1989) correlational information also provides information on the validity of teacher ratings, since researchers often depend heavily on assessments of achievement provided by teachers. Hoge and Coladarci (1989) reviewed the published literature on teacher judgments of student achievement and actual student performance on standardized tests. They included analysis of 16 studies where judgmental and test data were collected concurrently. In all studies reviewed, the accuracy of teacher judgments was assessed by examining the correlation between judgment and criterion. When the results of all 16 studies are taken together, they yielded a median correlation coefficient of \( r = .66 \). This suggests a moderate to strong relationship between teacher judgment
and student achievement, despite methodological differences in the studies. The authors concluded that there are "generally high levels of agreement between the judgmental measures and the standardized achievement test scores" (p. 308).

In two comprehensive empirical studies, Kellaghan, Madaus, and Airasian (1982) and Pedulla, Airasian and Madaus (1980) found significant correlations between teacher ratings and standardized achievement test scores. Even in cases where teachers may be unbiased by exposure to standardized tests, there are similarities in agreement between their ratings and test scores. For example, Kellaghan et al. (1982) performed an investigation of this issue in Ireland, involving 155 schools. Ireland was chosen as the test site due to the unfamiliarity of standardized testing procedures there; this was the first time tests of that nature were administered to school children in that country. The Drumcondra Attainment Test (1977) was used, which is a standardized, norm-referenced test developed for use in Ireland. Reading vocabulary, reading comprehension, and total math sections were analyzed. It was asked whether provision of test information affects subsequent teacher judgment of the students. There were two treatment groups, one in which
students were tested, but the results were not returned to the teachers, and a second group where tests were administered, and norm-referenced information was supplied to the teachers. It was hypothesized that initial teacher ratings were free from the influence of standardized tests, since they had never used these types of tests. A correlational analysis yielded moderate correlations between teacher ratings and test scores, with the highest ($r = .67$) between overall reading and teacher rating at the sixth grade. There was no significant difference between treatment groups when teacher ratings and test score correlations were examined. The authors concluded that the expectancy process operates within teachers with or without information on students' standardized test scores.

In addition, Pedulla, Airasian and Madaus (1980) analyzed teacher ratings of Irish students on group-administered IQ, math, and English tests. The ratings of 170 teachers on 2,617 students in fifth grade were factor analyzed. In this study it was also found that teacher judgments in all areas "tap a dimension similar to that tapped by standardized tests" (p.307). However, they also found that teacher judgments of academically-related behaviors, such as attention span and persistence, were also related to teacher judgments in math and English.
Standardized test scores were not highly related to attention span and persistence, and were therefore found to be more objective.

Similarly, Hopkins, George, and Williams (1985) found high correlations between teacher judgments and scores in reading ($r=.93$), language arts ($r=.95$) and math ($r=.91$) on the Comprehensive Test of Basic Skills (CTBS) (CTB/McGraw Hill, 1977). It was concluded by Hopkins et al. (1985) that "The test results, assessing the extent to which students are attaining curricular objectives and skills, are in relatively close agreement with evaluations teachers make after many hours of direct instruction and observation" (p. 181). Hopkins and Stanley (1981) also demonstrated that standardized achievement tests have moderately high correlations with school grades.

Kellaghan et al. (1982) also note there is a popular belief that teacher expectations are more likely to operate to the detriment of children of lower socioeconomic status than those from higher economic levels. However, in this study they concluded that "Little evidence that test information has a different impact on pupils from low SES than from other levels of SES" (p.197) was found. Therefore, overall, it was shown that teachers' perceptions showed considerable congruence
with test scores, and therefore are indicative of test validity.

In contrast to these results, also using low SES children, Manning, Manning and Cody (1985) compared the California Achievement Test (CAT) (CTB/McGraw Hill, 1977), the Alabama Basic Competency Test (ABCT) (Alabama State Department of Education, 1980) and informal reading measures: Houghton Mifflin Informal Reading Inventory (Houghton Mifflin, 1983), the Classroom Reading Inventory (CRI) (Silvaroli, 1984), a cloze procedure, and teacher judgment, on a sample of 58 Black, third grade, low-income students. Low income status was determined by participation in the free lunch program. High correlations were found between the instruments, the highest between the CAT reading and cloze reading ($r = .70$) and CAT reading and CRI ($r = .75$). Other correlations were in the moderate range, for instance when the norm-referenced CAT and criterion-referenced ABCT were compared, CAT total and ABCT vocabulary yielded a correlation coefficient of $r = .58$; and CAT total and ABCT comprehension correlated $r = .64$. Despite the moderate to high correlations, they reported that on the CAT, students scored much lower than on the ABCT. However, the authors did not report mean scores, nor did they perform any
statistical analyses to determine significant mean score differences, therefore it is not possible to ascertain whether students actually did score higher on informal and criterion measures than standardized measures.

Regardless of the reported lower scores on standardized tests by Manning et al. (1985) using the child’s lowest reading score (CAT), teacher judgments were consistently found to underestimate the child’s reading level. Based on the test results, 50% of the children were assigned to readers too easy for them, even using the lowest estimate of their reading performance. Therefore, it may be possible that teachers have lower expectations for children of low SES levels, and that standardized and informal tests actually reflect a more accurate picture of a student’s level of achievement.

Test Content and Curricular Mismatch

Another controversial area related to test validity in standardized testing is the match between test and curriculum, which some research suggests has an impact on test scores (Bianchini, 1978), and some does not (Mehrens & Phillips, 1986). Mehrens and Phillips (1986) conducted studies in two large school districts (populations of
23,000 and 36,000 students) with a sample of 3,380 students in one district in grades 3 and 6, and 3,300 in grades 3 and 6 from the second district. Within these districts, several different textbook series and a variety of curricula were in use. They found that neither the curricular match nor the textbooks used had an impact on standardized test scores. It was concluded that standardized tests are designed to measure broad, general, academic skills, and that generalized learning should be inferred from the tests. They further stated that it does not handicap students if they are taught general skills rather than specific objectives measured by the test.

Group-Administered Comprehension Tests and Students At-Risk

Few investigations with students defined as "at-risk" have been conducted. However, a search of the literature revealed two studies specifically addressing reading comprehension testing with students with below average achievement and low socioeconomic status. Carron (1978) and Powers and Gallas (1978) both found that students at-risk (defined by participation in a free lunch program) performed significantly lower on standardized, group-administered comprehension tests than students not
classified as at-risk. Carron found that when 500 students (270 Black and 230 White) in an urban school system were tested with the Diagnostic Reading Scales (Spache, 1963) Blacks scored lower than Whites, and poor Blacks scored lowest. He concluded, "As far as reading performance is concerned, it is bad to be poor; worse to be Black; still worse to be Black and poor" (p. 5). Carron did not have a substantial number of poor Whites in his sample (N=3), therefore it is not possible to determine how this group would perform in comparison to the poor Blacks.

Powers and Gallas (1978) tested 89 urban students in a Title I program designed for students at-risk in fourth, seventh, and ninth grades, using the comprehension subtest of the Comprehensive Test of Basic Skills (CTBS) (CTB/McGraw Hill, 1977). The students were administered the appropriate grade-level test with results indicating that about 50% of the students scored at or below the chance level for the test. However this study offered no insights into reasons for reading failure of these students.
Conclusion

When considering issues related to standardized testing, both with regard to the relationship between teacher ratings/test scores and test content/curricular mismatch, it seems that if tests are interpreted correctly and used appropriately, the information can be useful to the schools. According to Farr and Carey (1986), "If they are properly used, tests can be useful supplements to the information base needed for planning instruction, estimating students' reading development, and evaluating a school district's success in achieving stated goals" (p. 31). When group-administered tests are used in the schools, sensitivity toward appropriate use of results as well as assessment of reasons for students' lower scores is important, especially with an at-risk population. Given the paucity of research with standardized reading comprehension tests and the at-risk population, it is unclear what variables are acting to decrease their test scores.
Theories on the Nature of the Reading Process

Views on psychological processes implicated in reading have gone through a variety of changes in the past few decades. In regard to research performed to verify these theories, McNamara, Miller and Bransford (1991) state that "researchers have moved from a primary focus on memory for lists of words; to comprehension of, and memory for, meaningful sentences; and, finally, to the mental processes involved in understanding and remembering sets of sentences that form coherent texts" (p. 490). Several theories exist that propose disparate views regarding how this process occurs, however, most recent theorists suggest that reading is a complex, interactive cognitive process that involves many components often difficult to separate. A brief review of bottom-up reading processing models and top-down processing models is presented mainly for historical reasons. However, more emphasis is placed on recent models that are interactive in nature, since these theories are more congruent with the purpose of this study.
LaBerge and Samuels Model

LaBerge and Samuels (1974) posited a model of automaticity in processing. They attempted to identify components of the reading process, trace the path of this information and note changes in the form of the information as it passes from the printed page to deeper semantic and linguistic centers of the brain. Attention is a critical component for LaBerge and Samuels, since they theorize that comprehension cannot take place if attentional factors are focused on decoding. First, decoding occurs when featural information is analyzed and combined to form a word. Secondly, and as a result, comprehension is the synthesis and integration of information decoded, which requires attention. LaBerge and Samuels note that the more demanding the reading task, the more attention is needed to perform the process, but once a process is automatic, then no attention is required.

Goodman's Psycholinguistic Model

Goodman (1970) theorized the reading process begins with the intentions, predictions, and hypotheses of the reader. The reader selectively attends to only enough featural (visual) information to confirm or
reject her predictions. Goodman's model de-emphasizes the decoding of words, but instead concentrates on nonvisual information when reading. This information is used to hypothesize and predict, often through context, the message conveyed through print. The actual visual or graphophonic information of the printed page is used only to the extent to absolve any remaining uncertainty on the part of the reader.

**Interactive Models**

Recent theorists have moved toward an interactive perspective on the reading process. Research in reading as well as cognitive psychology has also moved in that direction. Stanovich (1991) has stated "newer conceptualizations are more data based than earlier frameworks" (p. 419). He also asserted that "more recent global models virtually all embody some type of hierarchical structure whereby the meanings activated by the successful recognition of words (the process of lexical access) are the building blocks for subsequent comprehension processes" (p. 419). Three major reading/cognitive processing theories are discussed in this section.
Rumelhart's model. A pioneer of the interactive theories, Rumelhart (1976) explained the reading process with a cognitive and sensory component. Rumelhart accounted for both readers' intentions, predictions and background knowledge as well as input from the printed page. Several assumptions are integral to Rumelhart's model: (a) letters are processed in terms of the surrounding letters (b) words are processed in terms of the surrounding semantic and syntactic features of the sentence (c) syntactic information is analyzed in terms of the semantics of the word string and (d) the passage is analyzed in terms of the overall context in which it occurs. Graphophonic information enters the reader's visual information store. A higher level feature analyzer then selects relevant featural information and provides it for the pattern synthesizer. Simultaneously, the pattern synthesizer is also receiving nonsensory information (while receiving sensory) from higher level areas with semantic, syntactic, orthographic and lexical information. These processes converge at the level of the pattern synthesizer which then selects and develops the most probable interpretation.
Stanovich's model. A more recent advancement, although similar to Rumelhart's model, is Stanovich's (1980) interactive-compensatory model. In addition to several of the assumptions of Rumelhart, Stanovich also includes a compensatory processing component. In his model of reading, Stanovich stated, "a deficit in any knowledge source results in a heavier reliance on other knowledge sources, regardless of their level in the processing hierarchy. Thus, according to the interactive-compensatory model, the poor reader who has deficient word analysis skills might possibly show a greater reliance on contextual factors" (p. 63). Stanovich asserted that first, one must assume bottom-up processing theories to be invalid. Stanovich indicated that knowledge sources at all levels contribute to pattern synthesis simultaneously. He asserted that rapid, context-free word identification allows capacity to be freed for comprehension processes. In the fluent reader, automatic processes are working (rather than conscious prediction) during reading, and it "short circuits" the conscious attention mechanism. However, a reader with poor context-free reading skills uses additional processes to aid word identification. Thus, this process uses
attentional capacity and leaves fewer cognitive resources for comprehension.

Stanovich (1991) emphasized the recurrence of research-related data that supports the contention that word recognition is the foundational process of reading. However, he stated that "to emphasize the centrality of word recognition is not to deny that the ultimate purpose of reading is comprehension" (p. 418). Stanovich pointed out the high incidence of people who display word recognition problems that have concomitant deficits in comprehension, where the reverse is almost never true. "Efficient word recognition seems to be a necessary but not sufficient condition for good comprehension" (Stanovich, 1991, p. 419).

Just and Carpenter's model. Recently, a model which involves a substantial amount of parallel processing, suggested by Just and Carpenter (1987), explains the reading process as a system which is comprised of a set of condition-action rules that substantially involves working memory. Their model is comprised of several recognize-act cycles, which can operate simultaneously. In the Just and Carpenter model, individual differences in the efficiency of word recognition are linked to differences in comprehension
processes. This model does not allow expectancy-based processing (hypothesis testing) to influence feature extraction from words, since that type of higher-order process is said to occur at the post-lexical level.

Therefore, most recent models have shifted to emphasize word recognition as a fundamental component of the reading process. Stanovich (1991) indicated that the capacity used by word recognition appears to decrease as reading skill increases, but even obligatory word recognition processes of fluent adult readers still utilizes some capacity. This is contrary to the assertions of the bottom-up theories that equate automaticity with capacity-free processing.

Other models. Additionally, more global, information-processing models that are interactive in nature include that of Luria (1966), Das, Kirby and Jarman (1979), and Naglieri (1989). These researchers and theorists seek to explain overall cognitive processing and many have conducted research on reading behavior and seek to explain it within the context of their model. Specifically, Naglieri and Das (1988) have elaborated, refined, and operationalized Luria's neuropsychological model of information integration,
and it has been termed PASS (Planning- Attention- Simultaneous- Successive). Most recently, when all components of the PASS model have been tested with reading disabled individuals, deficits in planning and attention, respectively were found to be a significant factor contributing to poor reading skill (Bardos, 1988). This shows that reading disabled children likely have difficulty controlling and regulating the coding processes appropriately, and possibly do not use attentional components constructively.

The issue of cognitive control and regulation is addressed in the literature on metacognition and reading, which is detailed in the following section.
Metacognition and Reading

With the increasing emphasis on the reader as information processor, concerns about the impact of awareness and regulation of these processes has led to research in metacognition. While cognition implies the actual processes and strategies used by readers (Forrest-Pressley & Waller, 1984)—for example, when a child remembers what was read, memory processes are used, and when words are identified decoding processes are used—metacognition refers to an individual’s ability to understand and manipulate his or her own cognitive processes (Reeve & Brown, 1984). One of the first to discuss metacognition was Flavell (1976) who stated,

Metacognition refers to one’s knowledge concerning one’s own cognitive processes and products. It refers to... the active monitoring and consequent regulation and orchestration of these processes in relation to cognitive objects or data on which they bear, usually in the service of some concrete goal or objective. (p.232)

Recently, there has been interest in investigation of metacognitive processes in studies of cognitive development and in reading research. These studies indicate that what children know about reading
strategies, goals, and other reading tasks can influence how well they monitor and plan in their own reading (Jacobs & Paris, 1987). According to Reeve and Brown (1984) and Baker and Brown (1984) central to the study of reading from a cognitive framework is the investigation of an individual’s problem-solving processes, not only the outcome of these activities. Since recent reading theorists view the individual as an active problem solver, examination of various strategies and plans used by a reader are of primary importance in reading research.

Baker and Brown (1984), Kirby (1984) and Flavell (1976) indicate that metacognition involves two types of activities: (a) knowledge of cognition and (b) regulation of cognition. Baker and Brown assert that knowledge of one’s cognition is important for reading. If one is aware of various strategies needed to perform effectively, then steps can be taken to meet the demands of the situation. Conversely, if one is not aware of strategies, it cannot be expected that the individual will take steps to prevent problems that may lead to comprehension failure.

Forrest-Pressley and Waller (1984) assert that metacognition involves conscious awareness of
processes, and a person must be able to communicate about them in some way. This differs from ideas of certain other theorists, for example, Das (1984), who indicated that individuals are not always able to state plans, decisions, and goals. He suggested that:

These [plans] need not be articulated, nor even break into the realm of awareness. We decide without always being able to justify our decision by verbal elaborations. Many plans indeed function like a hidden agenda determining an individual's activity. In fact I suspect that the higher forms of thinking are relatively inarticulate. We have to accept that although planful behavior can be subjected to psychological study and testing, all planful behavior may not be available for this purpose. (p. 47-48)

In agreement with Das (1984), Lawson (1984) believes that not all executive processes are conscious, controlled processes. He asserted that many processes are automatic, and are not reportable by the individual. However, this would not seem to indicate that all cognitive processes are inaccessible. Jacobs and Paris (1987) agree with Forrest-Pressley and Waller (1984) that for those cognitive processes where awareness is possible, the awareness indicates that cognitive states and processes can be shared, communicated, discussed, and made public. They believe that the concept of metacognitive strategy use
therefore, does not refer to higher level, inaccessible automatic processing. Jacobs and Paris (1987) stated, "By restricting the definition to shared knowledge we constrain the term considerably but avoid false inferences about the occurrence of metacognition when it cannot be observed or measured" (p. 258).

In addition to knowledge of cognition, one must consider the **regulation** of cognition. The self-regulatory function of metacognition includes checking outcomes, planning the next move, monitoring the effectiveness of the action, and testing, revising, and evaluating the strategies used (Baker & Brown, 1984). In this sense, metacognitive processes refer to the executive processes that direct cognitive processes and lead to efficient use of cognitive strategies (Forrest-Pressley & Waller, 1984). From this point of view skilled reading cannot consist only of knowledge of decoding and comprehension strategies, but also must include the ability to control these strategies. (Forrest-Pressley & Waller, 1984).

Jacobs and Paris (1987) reported that there have been relatively few studies that measure children's metacognition about reading. Paris, Wasik, and Van der Westhuizen (1988) reviewed the literature on this topic
found in refereed journals. They noted that prior to 1980, most studies did not measure metacognition, but rather investigated specific reading strategies per se. A few studies existing before 1980, though not specifying metacognition as a target of investigation, did present data providing insights on metacognitive activities and reading.

Paris et al. (1988) found that more than one-half of all published articles on metacognition and reading are discussion articles written for teachers. Few empirical studies in refereed journals were found to exist. They suggested that given the importance of the topic, more empirical research is needed to advance the field.
Factors Influencing Student Performance on Standardized Comprehension Tests

There are several factors that may impact standardized test performance, not related to the validity of the test itself, but to characteristics of the individual test taker. Taylor and White (1982) noted that research has been conducted with factors influencing individual testing performance on IQ tests (such as examiner effects), but little research has been conducted using multiple-choice, group-administered tests. Major factors that may affect group-administered reading tests can be logically divided into these categories: (a) specific reading problems of the individual, (b) strategic factors (such as test-taking strategies and reading strategies), and (c) mode of test-taking. Implications of these factors are addressed in this section.

Specific Reading Problems

Adequate reading consists of many facets, which may generally be divided into three broad components: (a) word recognition and word identification, (b) knowledge of word meanings, and (c) comprehension. A
reading problem in any of these three areas can affect comprehension test scores.

**Word recognition and word identification.** Word recognition is defined as ability to instantly read an individual word (McCormick, 1987). If a student lacks the necessary vocabulary to recognize common words, this can impede comprehension. According to McCormick, research has shown that readers need instant recognition of about 90% of the words in a text to read independently. If fewer than 90% are known, the flow of reading is interrupted in order for the reader to identify unknown words, which in turn, disrupts thought and concentration, often resulting in poor comprehension.

The term word identification, on the other hand, indicates a mediational component, that is using a strategy to decode unfamiliar words (McCormick, 1987). Three major word identification strategies are employed by skilled readers: (a) phonic analysis, (b) structural analysis, and (c) use of context clues. **Phonic analysis** is defined as use of individual letter sounds and groups of letter sounds to identify words. **Structural analysis** consists of breaking words into familiar parts, such as the base word and suffix, to
decode them. Use of context clues entails the use of surrounding words and their meanings to determine unknown words (McCormick, 1987). A student may be unaware of—or inappropriately or ineffectively use—one or more of these word identification strategies. Ignorance of these strategies, or their proper use, can lead to a breakdown in reading comprehension due to the inability to decode individual words while reading text or tests. As Johnston (1983) stated, "If children have to expend much of their cognitive capacity identifying unfamiliar words in the test, less remains for building a model of meaning" (p.31).

In a study that investigated factors influencing standardized reading comprehension test performance, Drum, Calfee, and Cook (1981) analyzed the characteristics of individual test questions related to low percentages of correct responding (as reported in the test manuals). They found the performance of children in grades 1 through 4 decreased as the percentage of unique content words increased—that is, as more words not included in the Dale-Chall list were introduced, performance declined. This indicates that students most frequently missed questions which included uncommon words. The researchers asserted that
an increase in unfamiliar words may lead to word recognition or word identification difficulties, which in turn, affect comprehension. Therefore, it appears that problems in areas of word recognition and word identification are likely to significantly influence performance on standardized comprehension tests.

**Meaning vocabulary.** One contributor to poor comprehension may be poor vocabulary knowledge and development (Anderson & Freebody, 1979). Anderson and Freebody (1979) asserted that if a student lacks familiarity with many word meanings, then it is likely that s/he will be a poor reader. Freebody and Anderson (1983) investigated the effects of increasingly difficult vocabulary on the reading comprehension of sixth-grade students. They found that by increasing the vocabulary difficulty, comprehension scores decreased. Beck and McKeown (1983) also found that knowledge of word meanings were related to comprehension. They taught word meanings to a group of fourth-grade students and found comprehension increased significantly for the group who received the vocabulary instruction. Similarly, Mason, Kniseley and Kendall (1979) studied the effects of polysemous words on sentence comprehension, and found that comprehension
declined when words with more than one meaning were introduced. However, they also suggested that this may indicate that the students did not effectively use context clues to determine word meaning while reading. In any case, knowledge of word meanings may affect comprehension test scores.

Comprehension. Although there are other factors influencing test performance, ideally a comprehension test should reflect only an individual's level of comprehension, that is, performance on the test should have veridicality with the test-taker's true comprehension level. There are two broad types of comprehension that could affect student performance: (a) literal comprehension and (b) higher-level comprehension. Literal comprehension is defined as ability to understand and recall details or facts directly stated in the text. Higher-level comprehension, on the other hand, involves interpretive thinking, and in some cases evaluative and creative responses (McCormick, 1987). Drawing inferences or conclusions from reading are examples of higher-level comprehension. Difficulty with either type of comprehension obviously can negatively influence test performance.
In addition, reading researchers have demonstrated the effect of background knowledge on comprehension with investigations of schema theory (Anderson, 1977). Higher level comprehension, and to some extent, literal comprehension is affected by prior knowledge. Anderson has stated that, "Knowledge a person already possesses has a potent influence on what he or she will learn and remember from exposure to discourse" (p. 2). Research by Johnston (1984b), Feeley and Wepner (1985) and Lipson (1983) consistently concluded that background information or prior knowledge has a significant effect on reading comprehension. For example, Feeley and Wepner (1985) tested comprehension of college students using a standardized test, with the experimental group receiving readings on topics included in the standardized test and the control group receiving readings on other topic areas. When students were retested with an alternate form of the test, the experimental group improved their scores significantly more than the controls, indicating the effect of prior knowledge in correctly answering test questions.

Johnston (1983) pointed out that lack of familiarity with either the content of the passage or the structure (narrative vs. expository) is an
indication of poor background knowledge for this text, which, in turn, can impede reading progress. Therefore, it is important to consider the impact of the text/background information mismatch or the inappropriate use of background information when analyzing errors made by students taking tests--since these may be viable sources of error significantly affecting comprehension.

However, according to Johnston (1984b) and Feeley and Wepner (1985), when assessing reading comprehension ability with standardized tests that are group-administered rather than individually-administered, it is not possible to know what part of an individual’s score is due to reading comprehension ability and what part is related to prior knowledge. Thus, on a standardized, group-administered test, a student with a paucity of background information on a topic, might perform poorly due to this factor, whereas another student may perform poorly due to inadequate reading comprehension skills (Feeley & Wepner, 1985; Johnston, 1984b). This is an important distinction, since both types of problems require different remediation efforts. However, if both students obtain the same score, it is impossible to distinguish the reasons
given the nature of a group-administered test. That is, there is no interaction between students and examiner as students select from predetermined multiple choices for answers, and respond simply by circling a response or blackening a circle on a computer analyzable answer sheet.

The problem of inadequate background knowledge or background/text mismatch also speaks to the cultural implications of reading comprehension testing. Researchers have emphasized the importance of culturally appropriate reading instructional materials and tests (Johnston, 1984b). Several studies have addressed the issue of cultural influences on standardized tests and have found variable results. Lipson (1983) found a background knowledge effect when investigating the ability of Catholic and Jewish children to recall information related to their religious instruction. On the other hand, Hochman (1973) administered two standardized reading tests, one in Standard English and one in Black vernacular dialect, and found that the children who performed poorly on one, performed poorly on the other, indicating no significant difference between tests related to dialect differences.
In contrast to Hochman's results, are those of Hall, Reder, and Cole (1975) who investigated verbal recall of children when they were read stories written in different dialects, either similar or different from their own. They found that Caucasian children performed better than Blacks in Standard English, Blacks performed better than Caucasians in Black English vernacular, and Blacks and Caucasians performed equally well when given their own dialects. The research by Hall et al. (1975) did not include administration of a reading test, but instead involved the recall of a story read aloud to the students. Although the implication remains that dialectical differences may impact performance on this language task with respect to prior information/cultural background, it is not clear from this study if the impact will generalize to standardized testing.

Desberg, Marsh, Schneider, and Duncan-Rose (1979) compared dialect effects of Black and Caucasian children on word recognition and reading comprehension subtests of a standardized test. They found that dialect did not interfere with reading comprehension or word recognition. The only significant difference between groups was that the Black dialect speakers
outperformed the Caucasian students on items in Black dialect, and were found to be bidialectical. Overall, empirical evidence does not support the contention that standardized tests in Standard English depress scores of Black English vernacular speakers (Desberg et al., 1979; Hochman, 1973), however, since these findings are generally based on group means there may be individual differences in this regard.

**Strategic Factors**

Two major factors that may influence reading comprehension test performance are reading strategies and test-wiseness (also known as test-taking strategies).

**Reading strategies.** Forrest-Pressley and Gillies (1983) assert that the effective use of reading strategies leads to successful, efficient reading. The mature reader has several different strategies available for use, depending on the nature of the reading task. For example, in a comprehension task the reader may skim, reread, or paraphrase depending on the demands of the situation. The skilled reader evaluates the situation and chooses the most efficient method available. Forrest-Pressley and Gillies assert that mature readers spontaneously use various strategies,
and actively monitor and adjust these strategies, and that these executive functions (monitoring, adjusting) are responsible for directing the flexible use of reading strategies.

Studies by Brown and Smiley (1977), Brown, Smiley and Lawton (1978) and Clay (1973), respectively, revealed that both young and poor readers do not use effective strategies for locating main ideas, remembering key elements of text, and monitoring meaning. Baker and Brown (1984) assert that many comprehension problems can be linked to students’ unfamiliarity with strategies or their appropriate use.

One of the first studies specifically investigating metacognition and reading was conducted by Myers and Paris (1978). They interviewed 40 second- and sixth-grade students about their use of reading strategies to investigate developmental trends of metacognition. Children’s knowledge of strategies was evaluated through the use of an individually-administered questionnaire. Awareness of strategies was measured by questions in areas of: rereading, inferences, imagery, and comprehension monitoring. To assess the regulatory function of metacognition, children were also asked questions pertaining to how
they resolved comprehension failure. Myers and Paris found that young children were unaware of the need to use different strategies for different materials and goals. They reported few strategies or reasons for checking understanding. In contrast, older children were aware of several strategies and could verbalize their appropriate use.

Similarly, Paris and Myers (1981) investigated perceived understanding of useful and harmful reading strategies, using 32 fourth-grade students. After reading a short text, subjects were verbally presented with 25 reading strategies: 10 negative, 10 positive, and 5 neutral. For example, a positive strategy was "imagine the story like a movie in your head" and a negative strategy was "say every word over and over." The children were asked to rate each strategy on a scale from 1 to 9 according to the utility of that strategy for them. Results indicated that poor readers were less aware of the detrimental influences of the negative strategies on comprehension than good readers; poor readers rated significantly more negative strategies as being positive. However, ratings of the positive and neutral factors did not differ between groups. In addition, the researchers compared the
amount of information recalled in a free recall to the amount of reported strategy use. They obtained a significant correlation \((r = .71)\) between number of strategies reported and the number of grammar units recalled. Therefore, strategy use was associated with amount of story recall.

Olshavsky (1976-1977) investigated reading strategy use in terms of three variables: (a) reader interest in the material--high versus low (b) reader proficiency--high versus low and (c) writing style of the author--concrete versus abstract. Using a modified protocol analysis procedure, 24 tenth-grade students' reading performances were analyzed and strategies inferred. Olshavsky concluded that readers with high interest and high proficiency use strategies more often. She also reported increased strategy usage with the abstract passage. However, upon closer scrutiny of her results, there actually were no statistically significant differences between any of the conditions.

Kavale and Schreiner (1979) investigated reading strategies used by good and poor readers as they responded to a standardized measure of reading comprehension developed by one of the experimenters. Also using a modified protocol analysis procedure, they
questioned 16 sixth-grade students, judged as having good verbal ability, about their reading strategies. The responses were classified into several categories. It was found that most often students' verbalizations concerning reasons for answer choices were related to information in the stimulus passage, and the information was synthesized by one of the following processes: (a) comparison--comparing the answer choice to information in the passage (b) classification--relating the answer choice to a major idea (c) definition--giving a definition of a word in the answer choice that related to the passage and (d) expansion--expanding on information from the passage in explaining the answer choice. Kavale and Schreiner found that above average readers evidenced greater flexibility and frequency of strategy use, whereas average readers applied strategies less frequently and less successfully.

Wade (1990) investigated reading comprehension using think alouds with a group of 52 students categorizing them into five types of processors: (a) good comprehenders- combination of top-down and bottom-up processing (b) non-risk taker -highly bottom-up (c) non-integrator- somewhat bottom up
(d) schema imposer- somewhat top-down (e) storyteller- highly top-down. She concluded that too much of either type of processing was not good, and that those who struck a balance between top-down and bottom-up processing were the most successful comprehenders.

Recently, Kletzien (1991) investigated strategies used by good and poor readers of expository text of varying levels. Forty-eight high school students were evaluated using verbal report techniques. Results indicated that good and poor comprehenders used the same strategies for each of the three reading levels. However, good comprehenders used a variety and persisted trying to use strategies even in the face of frustrating material. As passages became more difficult, differences between groups became more apparent. Good readers controlled strategy use and changed strategies depending on the passage difficulty. In contrast, poor comprehenders showed a decline in strategy use (in variety and number) as passage difficulty increased. Kletzien concluded "because the passages were of the same relative difficulty for both groups, and because both groups of readers appeared to be familiar with the same strategies (judging by
strategy use on other passages) one can conclude the difference between the groups was in regulation rather than knowledge of comprehension strategies" (p. 79). These results are consistent with those of Zabrucky and Ratner (1989) who also found support for the differential regulation of strategies between good and poor readers.

Pumfrey and Fletcher (1989) examined strategy use and oral reading miscues while subjects read texts of varying levels of difficulty. Using seven and eight year olds they found that as passage complexity increased, poor readers relied more heavily on graphophonic cues, and were unable to use semantic and syntactic cueing systems they used in easier passages. Similarly, Blaxall and Willows (1984) also found support for increased reliance on graphophonic cues as linguistic complexity of the text increased. The results of these studies support the interactive processing theories (e.g. Stanovich, 1980; Rumelhart, 1976; Britton, Glynn, Meyer, & Penland, 1982, Luria, 1980; Naglieri & Das, 1988), since the reader is able to display flexibility of strategies and compensate for difficulty in one cuing system by using another.
These studies have all emphasized the importance of reading strategy use, and its relationship to comprehension ability. It was consistently found that better readers seem more aware of strategies, use strategies more frequently, and use strategies more effectively.

Test-taking strategies. Test-taking strategies are a type of metacognitive/strategic factor that has been shown to impact a student’s test performance. Test-wiseness is defined as the use of the "characteristics and formats of the test and/or the test taking situation to receive a high score" (Millman, Bishop & Ebel, 1965, p. 710). This definition is the most widely used for the concept of test-wiseness. Test-wiseness principles include aspects of time-using strategies, error avoidance strategies, guessing strategies, and deductive reasoning strategies (Millman et al., 1965). Researchers have found that multiple-choice exams are especially susceptible to test-wiseness, and high correlations between measures of test-wiseness and high standardized test scores have been obtained (Alker, Carlson & Herman, 1969; Rowley, 1974; Sarnacki, 1979). Therefore, test-taking strategies were assessed in this
study, although it is difficult to separate test-taking and reading strategies when assessing student performance on a multiple-choice test (Farr et al., 1990).

In a pilot study concerned with test construction and item ambiguity, Haney and Scott (1980) questioned 11 second- and third-grade students about strategies they used to answer achievement test questions. The researchers analyzed ways children answered test items in an attempt to estimate the frequency with which the test questions were misinterpreted. They found that children use a wide range of test-taking strategies, and that a child's perception of item content often bears little resemblance to the test author's. For instance, the children sometimes marked a correct answer, but for an incorrect reason, and at times marked an incorrect choice, but displayed valid logic making the choice. Haney and Scott evaluated students' incorrect answers in terms of the following: (a) guessing, (b) elimination of wrong answers (c) reasoning (valid or invalid) (d) mistake (accidentally filling in the wrong choice or omission) (e) use of irrelevant extraneous knowledge, and (f) failure to use extraneous knowledge when needed. Results indicated
that in 91% of the cases, there was some direct form of reasoning or logic (valid or invalid) for selecting an answer. Guessing was the second most common means of choosing an answer.

Although not specifically addressed as a research question in their study, Haney and Scott (1980) observed that when children answered reading comprehension questions, "We did find instances in which misunderstanding of a single word in either reading passages or response alternatives led to children's selecting unkeyed answers" (p. 69). They suggested further research in the systematic analysis of children's accounts of how they perceive and reason about test items. Although students did not initially read the test questions and answers aloud to the experimenters in their study, for future research it was suggested that questions and answers be read aloud to more fully determine the effect of vocabulary knowledge on reading test performance. One difficulty with interpretation of Haney and Scott's results, is the classification of all mistakes as erroneous test-taking strategies, when they may also be related to other factors such as background knowledge/text mismatch or specific reading problems (e.g., word
identification strategies). Furthermore, a weakness of this study was that subjects were questioned the following day about their test-taking strategies. Although the experimenters questioned the subjects on individual test items, the time lapse between test-taking and questioning may have affected the students' ability to remember their strategies. The young age of the students could also have affected the results, since young children may have greater difficulty articulating their strategies.

A study investigating strategies used spontaneously to answer standardized reading test items and the effectiveness of these strategies was performed by Scruggs, Bennion and Lifson (1985). These researchers created a test consisting of several questions from the Stanford Achievement Tests (Madden, Gardner, Rudman, Karlsen, & Merwin, 1973) in the areas of word reading, reading comprehension, word study skills, and vocabulary. In the area of comprehension, they collected information on students' attention to distractors and referral to the reading passage to answer questions. Based on students' use, Scruggs et al. created five categories of strategies: (a) skipping, (b) procedural error, (c) guessing,
(d) deliberate strategy, and (e) knowing the answer. Results showed that subjects rarely referred to the reading passage; for example, in 89% of the cases where answers were incorrect, students had not looked back to the passage to either verify or search for an answer. Carelessness also was observed when students did not attend to distractors; in 40% of the cases, students had not read all the distractors when they had answered incorrectly. They found that "a complete hierarchy of test-taking skills exists beyond simply knowing or not knowing the answer, and they may be more or less effective on a standardized reading test" (p. 481).

The researchers also concluded that word reading errors had only a small effect on test performance. Scruggs et al. stated that, "Although performance was clearly impaired when students misread a word of key importance, in general misreading words was less detrimental than might be expected" (p. 483). They found that students still answered an average of 58% of the items correctly when one or more words were misread. Although Scruggs et al. indicated that this percentage was higher than expected, it seems this might be of concern in future research. In the Scruggs et al. analysis students who misread one word were not
separated from those who misread several words, and those who missed only one word, may have been more successful identifying the correct answer than those who had difficulty identifying several words. Therefore, Scruggs et al. failed to separate the effect of word identification difficulties from comprehension difficulties.

Farr et al. (1990) examined what happens when students take a multiple-choice reading comprehension test. Twenty-six college students were tested on several reading selections using either an introspective interview or a retrospective interview. They found that most subjects were driven by the test questions to find an answer. Test questions were often used to direct a search of the passage to find an answer. The goal of the students was not to read and comprehend the passage itself then demonstrate knowledge by answering questions, the questions themselves were intertwined in the reading task, and the goal became one of question answering. The most popular question-answering strategy was that of looking back to the passage to locate an answer, and reading all the answer choices and eliminating the wrong answers. Therefore, reading strategies and test-taking
strategies were interrelated and were found to be an essential aspect of reading when taking a multiple-choice test. Farr et al. conclude that multiple-choice comprehension tests seem to be a special type of reading task, and they are valid for at least one type of reading, a type which is common in school. Farr et al. suggest further research is needed with varying age levels and reading levels.

These studies, taken together, imply that test-taking strategies and reading strategies together have a significant impact on a student's performance on a standardized, multiple-choice comprehension test, and may be significant contributing factors to poor test performance.

Mode of Test-Taking

Another major factor that can impact student performance on a test is the mode of reading (silent vs. oral) and mode of test administration (individual vs. group). Although these factors are not specifically addressed in the present study, this study involved individual test administration with oral reading, and the subjects were selected by group test
administration with silent reading. Differences between these factors are briefly acknowledged.

**Mode of reading.** Although group-administered comprehension tests involve silent reading, differences may arise between silent and oral reading performances.

According to Holmes (1985), whether comprehension is differentially influenced by oral and silent reading is a controversial issue. A number of studies have addressed this topic. Some research supports the contention that comprehension is superior after oral reading (Mead, 1915, 1917; Pintner, 1913), some indicates that comprehension is better after silent reading (Collins, 1961; Duffy & Durrell, 1935-36; Elgart, 1978; Rowell, 1976), and some has shown equal comprehension effects after either mode of reading (Holmes, 1985; Juel & Holmes, 1981; Poulton & Brown, 1967; Rogers, 1937). According to Holmes (1985), "Although considerable controversy has been generated, research has not established clearly the superior mode for understanding or remembering" (p. 576). Therefore, mode of reading may impact a student's performance on a standardized comprehension test, although it is unclear at this time in what way it will be affected.
Group versus individual administration. According to Anastasi (1988) in group testing situations, as opposed to individual testing situations, the examiner is less involved with the students, and has less opportunity to establish rapport, obtain cooperation, and maintain the interest of examinees. She stated, "Any temporary condition of the examinee, such as illness, fatigue, worry, or anxiety that may interfere with test performance is less readily detected in a group than in individual testing" (p. 313). Anastasi further stated, "Persons unaccustomed to testing may be somewhat more handicapped on group than on individual tests" (p. 313). Willis (1970) found that for a sample of emotionally disturbed children, their performance was significantly higher on an individual test than a group test. Therefore, it is likely that motivation may impact a student's performance on a group comprehension test more so than on an individually-administered test.

Taylor and White (1982) investigated the impact of reinforcement on group-administered test scores. They hypothesized that students performed poorly due to lack of motivation toward the test. Students were given monetary reinforcement for improvement in test scores
on an alternate form of a standardized test. Results indicated that reinforced students obtained significantly higher test scores than students not reinforced. Taylor and White concluded that in normal testing conditions, "Students may not be motivated to try their best, but when students are reinforced for test performance, they tend to obtain test scores that are more indicative of their skills and knowledge than scores under non-reinforcing conditions" (p. 206). This research suggests that motivational level does have an impact on student's test performance.

Similarly, Lahey, McNees, and Brown (1973) investigated the impact of reinforcement on test performance with two children who tested below average in reading comprehension. They used a single-subject design to test the effect of giving pennies and social praise for correct responses on the Spache Diagnostic Reading Scales (Spache, 1963). Two subjects were tested, with two control subjects matched with the others on age, grade level, socioeconomic status, and oral reading level, however, the controls obtained average scores on the reading comprehension test. Therefore, the major differences between the two control children and the other two was level of reading
comprehension. There were two intervention phases and two baseline phases in the study. The results indicated an improvement in the percentage of correct answers in each of the intervention conditions with the two below average students, whereas no differences for the two control students were found when reinforcement and praise was initiated. The authors concluded that reinforcement may have a differential effect for low-achieving students. However, they encourage more research in this area, due to the limited generalizability of the testing situation.
Research-Related Methodological Issues

According to Jacobs and Paris (1987) most studies that have assessed children's knowledge about metacognition/strategy use have employed some form of individual interview or individual verbal report data. Kirby (1984) stated that unlike memory and cognition research where strategies are usually inferred from patterns of performance, metacognitive research tends to rely on individual interviews with subjects and are concerned with their verbal reports of knowledge and strategies. Kirby stated, "This self-report basis of data is probably the weakest feature of metacognitive research, although arguably a necessary one as long as awareness is the crucial characteristic." (p.55). According to Garner (1982) and Ericsson and Simon (1980) some difficulties may arise when experimenters give cues to particular strategic behavior. For instance, directed probes may suggest certain strategies to the subject, and their responses therefore simply reflect examiner expectations. However, according to Lawson (1984) verbal report studies elucidate the complexity of the executive processing involved in problem-solving situations and
he encouraged future research and refining of the interview technique in the study of executive processing.

According to Garner (1982), much energy has been spent on examining products such as recall protocols and responses to questions regarding what readers apparently did when reading. Garner questioned why so little time has been spent asking readers to report introspectively on what they did as they worked through a task. The reluctance to accept introspection as a method of scientific observation is likely the answer.

Several criticisms have been levied at the reliability and validity of verbal report data. According to Nisbett and Wilson (1977) people appear to have little ability to report cognitive processes accurately. They document several social and experimental psychology studies that have attempted (unsuccessfully) to use verbal reports as the main source of data. Garner (1987) reported eight criticisms of verbal report data on children's reading which were paraphrased by Jacobs and Paris (1987) in the following manner:
1. Some cognitive knowledge and processes are tacit and inaccessible, yet reflection on concurrent cognitive processing can often be accurate.

2. Verbal report data may include participants' rationalizations, elicited mimicry and fabrications because of the demand characteristics of the situation.

3. Children may lack the language and verbal facility to discuss mental events.

4. The behavior and characteristics of the interviewer can elicit answers perceived to be socially desirable.

5. Reliability of interviews is rarely assessed.

6. Forgetting may interfere with introspective reports.

7. Children have difficulty answering questions about hypothetical situations.

8. Asking questions during or after cognitive processing can disrupt thinking. (p. 264)
Despite these apparent weaknesses with verbal report data, Ericsson and Simon (1980), Farr et al. (1990), Wade (1990), and Kletzien (1991), argue that if such data is appropriately collected, it is an important source of information about cognitive processes. One reason for perceived problems with verbal report data, is that some authors have failed to distinguish between types. Ericsson and Simon draw a distinction between three types of verbal reports: (a) thinking aloud protocols, (b) retrospective responses to specific probes, and (c) classical introspective reports of trained observers. Thinking aloud protocols require the individual to think aloud as s/he solves a problem (Olshavsky, 1976-1977; Wade, 1990). Ericsson and Simon (1980) term this type of analysis "concurrent verbalization." The retrospective method involves reading an entire selection, then the subject responding to the total event by remembering what he or she was thinking and doing while reading (Olshavsky, 1976-1977). Ericsson and Simon note that this method can be effective if followed by general probes that follow the retrospective process. In the case of reading, Kavale and Schreiner (1979) noted, "The requirements of reading research, however, make it
necessary to modify the procedures of protocol analysis. Since subjects must read before they think aloud, the protocol analysis technique becomes partially retrospective" (p. 106). Introspection requires the subject to verbally report his processing as it occurs, for example to theorize about cognitive processes (Olshavsky, 1976-1977). This technique is not widely used, and it appears difficult and cumbersome in most cases. Introspection is often confused with thinking aloud protocols and retrospective reporting.

Often a combination of protocol analysis and retrospective techniques are used to analyze reading processes (e.g., Olshavsky, 1976-1977; Kavale and Schreiner, 1979; Farr et al., 1990; Wade, 1990). However, some researchers have employed retrospective techniques alone (e.g., Fareed, 1971; Garner, 1982; Smith, 1967) and concluded they were useful in reading research. For example, Garner (1982) used the method of retrospective reporting with 20 college students given a passage to read and summarize. Two conditions, delayed and immediate self-report were used, along with a student observer for each participant who recorded any observable behaviors during reading and
summarizing. No cuing of strategic behaviors was given. The results indicated that more thorough reports were given by students in the immediate self-report condition. Students were able to report several strategies used, and the author reported that the students listed all of the major summarization strategies listed in relevant literature. They also found that observations concurred with the reported strategies used most often in the immediate self-report condition. This finding denotes the importance of minimal delay between the task presentation and retrospective report response time.

When interviewing or obtaining retrospective verbal report data from a child regarding strategy use and errors, it is obviously important to obtain accurate results. A study by Steffenson and Guthrie (1980) revealed that in formal interviewing sessions, children performed in a different manner than when placed in an informal, more realistic situation.

Results of this study suggest that it is important to take the situational context into consideration when interviewing. They concluded that if the experimenter conducts the interview in a less formal manner, resembling a dialogue between the experimenter and
student, it is likely that more valid results may be obtained.

Therefore, use of verbal reports to assess cognitive processes, although controversial, is widely used and accepted as one method of collecting information about the reading process. Kletzien (1991) concluded, "it seems that verbal reports, although not perfect provide more complete information than can be obtained through observation or performance scores alone" (p. 71).

Summary of the Literature Review

Several issues were discussed in this review in regard to standardized testing. In terms of test validity, tests are valid only if used in a manner consistent with their intended purpose. There is controversy over the definition of reading comprehension and its measurement, and standardized, group-administered tests were frequently used to obtain data on students' reading levels. Reading test validity often has been challenged, although a substantial amount of research has provided evidence for criterion-related validity, with teacher judgment and curriculum/test match as criteria. Mixed results
have been obtained between teacher judgment and standardized test ratings with students from low socioeconomic levels. Research with at-risk students has shown that they score significantly lower than others on standardized, group-administered tests. Care should be taken to insure appropriate test interpretation, since if used correctly, they can provide useful information about students and schools.

Several reading theories were discussed and the shift from sequential, bottom-up and top-down models to more interactive, information processing models was presented. A re-emphasis on the importance of word recognition as a fundamental reading process was presented, although this does not negate the fact that comprehension is the main purpose of reading.

Several factors may impact students' performances on a standardized, group-administered comprehension test. A student may exhibit specific reading problems in word recognition, word identification, or comprehension. However, specific reading factors are not the only influential variables, although ideally, a comprehension test should measure only comprehension. Strategic factors such as metacognition and test-taking strategies can also enhance or depress test scores. A
student's level of motivation during testing obviously
has an impact on test performance, especially during a
group-administered test, since the examiner is not able
to interact with the students throughout the assessment
session. The analysis of each of these factors
impacting test performance is important to determine
their effects on an individual's test performance.

When assessing student test performance
individually, it is not only important to analyze the
answer (product), but also the individual's reasons for
a particular answer choice. This can lead to insights
into why a student made a particular type of error.
This type of analysis often involves gathering
individual verbal report data, using either protocol
analysis, retrospective reports, or a combination of
both. The purpose of gaining strategic information
necessitates the use of verbal report techniques as the
primary source of data gathering, although the
importance of accompanied observation is also noted.
Verbal report techniques have been criticized on
several grounds, however, when interviews are conducted
properly, verbal reports have been shown to provide
valuable information on a person's problem-solving
strategies. Knowledge of cognitive processing
strategies can provide information that may support or refute various theories of reading.

Research Questions

The following research questions were addressed in this study. The questions are divided into three areas, (a) errors between groups (poor readers versus average readers), (b) strategy use between groups, (c) strategy use within groups and (d) answer changes.

Errors Between Groups: Poor Readers versus Average Readers

To investigate group differences in error types, several questions which made group comparisons were asked.

Question #1 "What is the relationship between error types and reading ability, that is, do the error profiles differ between groups?" This question is specifically addressed in the following hypotheses.
Hypothesis #1: It is hypothesized that there will be a statistically significant difference between the number of word recognition/word identification errors made by poor readers when compared to average readers.

Hypothesis #2: It is hypothesized that there will be a statistically significant difference between the number of comprehension errors made by poor readers students when compared to average readers.

Hypothesis #3: It is hypothesized that there will be a statistically significant difference between the number of meaning vocabulary errors made by poor readers students when compared to average readers.

Strategy Use Between Groups: Poor Readers versus Average Readers

To answer the questions "Do poor readers differ from average readers in use of strategies?" Question 2 addresses comparisons of the profiles of the two groups.
Question #2: Do the profiles of strategy use differ between groups, that is, which strategies are associated with poor readers versus good readers regardless of the correctness of the response?

Strategies:

Strategy 1  Read all possible answer choices.
Strategy 2  Referred to passage to find answer.
Strategy 3  Read all answer choices and looked back to the passage.
Strategy 4  Read all answer choices, but then looked back and guessed or matched words.
Strategy 5  Read all answer choices and used a process of elimination to decide an answer.
Strategy 6  Immediate answer selection- no other strategy used.
Strategy 7  Looked back to passage and matched words, or guessed or matched words alone.
Question #3: As an indicator of effective strategy use, in what way do the profiles of poor readers and average readers differ when students obtain a correct answer?

Question #4: As an indicator of ineffective strategy use, what profile differences exist between groups when students select an incorrect answer?

Answer Changes During Individual Test Administration

Group differences on the propensity to change answers was compared.

Question #5: Is there a difference between poor readers and average readers when the frequency of answer changes is compared? This question is specifically addressed in the following hypotheses.

Hypothesis #1: There will be no significant difference between poor readers and average readers when overall number of answer changes are compared.

Hypothesis #2: There will be no significant difference between poor readers and average readers when answers were changed from incorrect to correct.
Hypothesis #3: There will be no significant difference between poor readers and average readers when answers were changed from incorrect to incorrect.

Hypothesis #4: There will be no significant difference between poor readers and average readers when answers were changed from correct to incorrect.
CHAPTER III

METHOD

Chapter Three provides a description of the subjects participating in the study, the setting for the study, the experimenter, the instruments and materials used, data analysis procedures, definition of variables, method of assessing reliability of the data, and the statistical analyses employed.

Subjects

Sixty sixth-grade students participated in this study, 30 students classified as high risk, and 30 classified as low risk. Sixteen subjects were evaluated during the summer (July and August) before they entered sixth grade; the remaining 44 students were tested in the fall and winter of their sixth-grade year. Subjects were selected from two types of locations, one, an urban area with a high concentration of high-risk students, and the other, two middle class suburban areas mainly comprised of low-risk students. Information regarding all locations is described thoroughly in a subsequent section.
To select the low-risk group (average readers), consent forms were sent to parents of the sixth-grade students in the two suburban school districts (see Appendix A). The experimenter administered, in a group setting, a norm-referenced reading comprehension test, the Gates-MacGinitie Reading Tests, comprehension section, Level 5-6, Form L (MacGinitie & MacGinitie, 1989), to three classrooms of students totaling approximately 75 students. This group assessment was necessary since these students did not have recent group-administered test data in their school records. The test was administered in the manner consistent with standardized procedures stated in the examiner's manual. Two-thirds of the average readers were selected in this manner (n=20). However, due to school policy in one district, the examiner was required to use existing test data for selection of the remaining one-third of the average students (n=10). On the Gates-MacGinitie Reading Test, students marked answers directly in the test booklet, by blackening circles provided, rather than using separate answer sheets. Marking answers in the test booklet is one of two methods for answer specification suggested in the test manual. The tests were scored and individual percentile ranks obtained, and the latter converted to standard scores. Students who
obtained average scores (standard score of 90 to 110; 25th to 75th percentile) on the Gates-MacGinitie reading test or the test data obtained from school records (Comprehensive Tests of Basic Skills, Comprehension section only; CTBS/ McGraw Hill, 1977). were separated from others. From these selected students, only those who were not participating in the free lunch program, (indicating non-poverty level) and who had parental consent, were selected as participants.

The final sample of low-risk students (average readers) consisted of 30 sixth-grade students, 17 boys and 13 girls, with a mean age of 11 years 11 months (range=11-1 to 12-11; standard deviation=6 months). There were 6 Blacks and 24 Caucasians, with IQ scores that ranged from 84 to 112 (mean=95.57; SD=7.79). Reading achievement standard scores ranged from 91 to 110 (mean=102.3; SD= 5.39). This indicates that students were average in both intellectual ability and achievement.

Of the 30 students in the average reader sample, 6 students lived in single-parent homes; 24 lived with both parents. Education levels of both the male and female guardians fell into categories 2, 3, 4, and 5 (see Appendix A) with the majority in category 3 and 4 (Category 3= High School Diploma; Category 4= 1 to 3 years
of college or technical school). For parental occupation, both male and female guardians' highest percentages were category 6 (operator, fabricator, laborer); the second highest percentage differed for males and females, 25% of the males fell into category 5 (precision production, craft, repair) and 20% of the females fell into category 2 (technical, sales, administrative support).

A similar selection procedure occurred for high-risk students (poor readers), with the exception that all 30 students were selected from a district-administered group test. This difference in selection procedure was necessary for the approval of this study in the urban school district. Four urban middle schools with large proportions of low socioeconomic status families were contacted for participation in this study. One school was willing to participate. Students were selected from the recently-administered Comprehensive Test of Basic Skills (CTBS; CTB/McGraw Hill, 1977) administered to all students in the spring of 1990 and obtained from the school district's Department of Evaluation Services. Students who scored at or below the 23rd percentile (standard score 89) on the reading comprehension area of this test were selected initially. At this time, consent forms were sent to parents for student participation in individual
testing. All students in the high-risk sample participated in the free lunch program. Eligibility for free lunch is based on low family income (as reported by the school). All students who returned parent consent forms were tested individually.

The final sample of high-risk students consisted of 30 sixth-grade students, 16 boys and 14 girls, with a mean age of 12 years 2 months (range=11-3 to 13-11; SD=7 months). There were 5 Blacks and 25 Caucasians, with IQ scores that ranged from 86 to 108 (mean=93; SD=5.75). Reading achievement standard scores ranged from 64 to 89 (mean=84.4; SD=6.72). This indicates that students were average in intellectual ability, but below average in reading achievement.

Of the 30 students in the high-risk sample, 11 students lived in single-parent homes; 19 lived with both parents. Education levels of both the male and female guardians fell primarily into categories 2 and 3, with only four exceptions (see Appendix B) (Category 2= 9th - 11th grade; Category 3= High School Diploma). For parental occupation, male guardians' highest percentages were category 5 (precision production, craft, repair), 6 (operator, fabricator, laborer), and 8 (not currently in labor force). Female guardians' highest percentage was
category 6 (operator, fabricator, laborer), followed by 7 (homemaker) and 8 (not currently in labor force).

When the final two groups were compared for descriptive purposes, there was no significant difference between IQ levels of group 1 (M=95.57) and group 2 (M=93.0), t(58) = 1.45, p>.05. Similar percentages of sex and race were also found (5 blacks versus 6 blacks, and 17 boys and 13 girls versus 16 boys and 14 girls). However, when age was compared, there was a statistically significant difference between group 1 (low-risk M=142 months) and group 2 (high risk M=146 months), t(29) = 2.43, p>.05). The high-risk students were slightly older than the low-risk students by an average of four months. The samples also differed in overall test scores, as planned. Mean test selection scores of group 1 (M=102) were significantly higher than group 2 (M=84), t(58) =11.38, p<.01). The average number of errors during individual administration of the reading test also differed significantly between group 1 (raw score errors M=12.83) and group 2 (raw score errors M=24.93), t(58) =11.74, p<.01). Socioeconomic status also differed between groups, group 2 had lower levels of parent education and income than group 1.
Setting

The data were collected in three locations. The location of the school of the high-risk students (location 1) was in a large midwestern city school system located in central Ohio. Specific district information regarding school size and composition was obtained. This district was comprised of 111 schools, with a total student enrollment of 64,057. Minority students make up 49.7% of the total student body, with a total of 34.33% who received Aid to Dependent Children (ADC). Total per-pupil expenditure was reported as $4,789, and the district average household income was $23,278.

The middle school from which students participated in this study was comprised of a total of 754 students, 287 in the sixth grade. For the entire school, 71.2% were non-Black, 28.8 were Black, with a percentage of males to females of 49.9% to 50.1%. In the sixth grade only, 71.4% were Caucasian and 28.6% were Black. Of the school personnel (48 certificated staff) 30% were Black and 70% Caucasian, with a total educational personnel to student ratio of 20.94 to 1. The school reported a total of 82.7% of students qualified for free or reduced price lunch, as compared to a district average of 56.3%. The mobility rate for the school was 45.9%.
Low-risk students were tested in two separate suburban areas in central Ohio. Location 2 was a district comprised of 22 schools with a total enrollment of 16,441 students, 6.93% minority enrollment, and with 32% on ADC. Total per-pupil expenditure was reported as $3,648, with an average family income of $25,113. The school was located in a somewhat rural area, although the district is considered suburban. The school was comprised of a total of 827 Caucasian, 43 Black, 4 Hispanic, 6 Asian, and 3 American Indian students, with a total of 53% males and 47% females. The teacher to pupil ratio was 1 to 29, with 100% of the certificated staff Caucasian. The school reported that 15% of their students qualify for free lunch, with 7% eligible for reduced price lunch. The mobility rate was reported as 21%.

Location 3 was a smaller suburban district with six schools, and a total student enrollment of 4,684. The total minority enrollment in this district was 8.13%, with 2.2% of students on ADC. The average family income was reported to be $28,309, with district per-pupil expenditures of $3,513. The school was comprised of a total of 620 students, of which 42 were Black. The teacher to pupil ratio was reported to be 1 to 25, with a
mobility rate of 27%. Of the total student body, 60 students qualified for free or reduced price lunch.

Group testing occurred in separate classrooms suitable for group test administration. Individual testing was performed in a quiet room designated for testing by the building principal.

**Experimenter**

A doctoral student in school psychology with a minor cognate area in reading served as the experimenter for the present study. The experimenter was a certified school psychologist experienced in IQ assessment with children as well as with reading assessment. She also had experience in direct supervision of reading disabled students in a university-based psychoeducational clinic.

**Instruments and Materials**

**Intelligence Test**

Matrix Analogies Test-Expanded Form (MAT-EF). The Matrix Analogies Test-Expanded Form (Naglieri, 1985) is an individually-administered test composed of 64 abstract designs of the Standard Progressive Matrix Type, printed one per page. The matrices are printed in blue, yellow,
black, and white to reduce the influence of impaired color vision. The test involves minimal motor movement and minimal verbal comprehension and is divided into four Item Groups, Pattern Completion, Reasoning by Analogy, Serial Reasoning, and Spatial Visualization. The MAT-EF yields a Total Test Standard Score with a mean of 100 and standard deviation of 15, and Item Group Standard scores with a mean of 10 and standard deviation of 3.

The MAT-EF manual reports that standardization took place in two phases. Phase 1 consisted of administration of 34 of 64 items to a sample of 4,468 students in class-size groups. Phase 2 consisted of individual administration of all 64 items to a sample of 1250 students. This allowed utilization of a larger sample that enabled more stable and precise estimates of the distribution of scores at each age interval. Subjects in both phases were between ages 5 and 17 and were representative of the United States on the basis of sex, age, geographic region, ethnic group, community size, and socioeconomic status. Reliability information reported in the manual consisted of internal consistency coefficients ranging from .88 (age 5) to .95 (age 6) for the Total Test, the median coefficient is .93. Test-retest
reliability coefficients range from .40 to .67 on the Item Groups and the coefficient for Total Test scores is .77.

Developmental changes in mean scores were cited as evidence of construct validity for the MAT-EF and a correlation of .70 was found between raw scores and age. Correlations between .58 to .64 were found for each MAT-EF Item Group raw score and age. The MAT-EF has high loadings on the general ability factor, and no sex differences nor race (Black-White) differences in mean scores were found. Criterion-related validity is reported in the manual. The correlations between the MAT-EF and the WISC-R (Wechsler, 1974) are Verbal IQ = .37, Performance IQ = .41, and Full Scale IQ = .52. These are all significant at the .001 level. The correlation found between the MAT-EF and the Raven's Coloured Progressive Matrices (CPM) (Raven, 1956) raw scores is .71 on a sample of normal children. The MAT-EF correlated significantly with the WISC-R Performance IQ (.143) and with Raven’s CPM (.43) on a sample of Native American children. For further information on the MAT-EF, see Naglieri (1985).

Reading Test

Form L was used for group administration to determine subject selection. Form K was individually administered to assess error types and strategy use, as described below. This test was a standardized, norm-referenced, reading test composed of two sections: Comprehension and Vocabulary. In the present study, only the comprehension section was administered. The comprehension section was designed to measure students' abilities to read and understand narrative and expository text (MacGinitie & MacGinitie, 1989). Level 5/6 was composed of 14 reading passages each followed by two, three, four, or five questions, yielding a total of 48 reading comprehension items. On this level of the test (for both Form K and Form L), there were eight narrative passages and six passages of exposition. One narrative selection was a poem. On Form K, there were two italicized introductory statements preceding two reading passages to provide background information to the reader regarding the context of the test passage. In each case, without reading the brief italicized sentence, it was difficult for the student to determine the meaning of the passage. In addition, the poem had a title that should be read by the student, but the other passages do not have titles. Each question following the passages was accompanied by four
answer choices where students were required to blacken with a pencil the corresponding circle in the test booklet. The test was administered according to group standardization procedures, and the students were allowed 35 minutes to complete the comprehension section. Total comprehension raw scores of all students were converted to standard scores, with a mean of 100 and standard deviation of 15.

The Gates-MacGinitie was standardized in October 1987 and April 1988. Norms for alternate forms K and L were based on the 1980 U. S. Census, using school districts stratified on the basis of geographic region, district enrollment size, and district socioeconomic characteristics. The sample included 77,413 students from kindergarten through 12th grade. The students were tested twice with alternate forms of the Third Edition of the test. Readability was assessed using several readability formulae, and content validity was assessed using a nationwide field test of several items, expert judgment, and consultation with minority group experts.

Reliability coefficients for the Gates-MacGinitie Reading Tests for internal consistency and alternate forms reliability were reported in the technical manual. Internal consistency reliability information was obtained
from both fall and spring standardization samples. For the comprehension section only, a Kuder-Richardson Formula 20 (KR-20) reliability coefficient of $r=.92$ was obtained for both fall and spring administrations of Level 5/6, Form K. Similarly, Level 5/6, Form L, yielded a coefficient of $r=.91$ for both administrations.

An alternate forms reliability coefficient of $r=.83$ was obtained when Form K and Form L were compared using a sample of 886 students. When raw scores between spring and fall testing were compared using 2,844 students (Form K, comprehension section only) a reliability coefficient of $r=.82$ was reported.

Comparisons between the Gates-MacGinitie Reading Tests and other tests yielded relatively high correlations. When Level 5/6 was compared with the California Achievement Test (CTB/McGraw Hill; 1977) on a sample of 50 students (comprehension section only), a correlation of $r=.81$ was obtained, when compared with the comprehension section of the Metropolitan Achievement Tests ($N=28$) a correlation of $r=.82$ was obtained, and when compared with the comprehension section of the Survey of Basic Skills, ($N=51$) a correlation of $r=.75$ was found.

The Gates-MacGinitie Reading Tests was also compared to reading course grades and grade point average, which
yielded moderate correlations. For sixth-grade students, the median correlation coefficient was $r = 0.60$ when course grades and the comprehension section of the test were compared. When total grade point average was compared to the comprehension score, the median coefficient was $r = 0.62$.

Graham (1990) in a favorable test review, concluded although the reading passages are short (most less than 200 words) they seem to have an adequate amount of narrative for comprehension. She also reported the comprehension questions themselves seemed unambiguous. The author cautions that the test is not a diagnostic test, but appears to be a valid measure of assessing general reading achievement.

If further information regarding the details of test development, item selection, standardization, and statistical characteristics is needed, extensive information is presented in the technical manual (see MacGinitie & MacGinitie, 1989, Technical Report).
Student and Experimenter Materials

Students were provided with appropriate test booklets and pencils. The experimenter used a tape recorder, a prepared script to explain procedures to the students, an Individual Observation Sheet and Strategy Checklist to note behaviors during testing (see Appendix B), and an Error Coding Sheet to categorize errors during the analysis phase (see Appendix C). To assess intellectual ability, materials for administration of the MAT-EF were used by the experimenter.

Data Collection Procedures

Individual Administration of the Intelligence Test

For those students who had met the standardized test criteria, socioeconomic criteria, and had returned parental consent forms, the Matrix Analogies Test-Expanded Form was individually administered to obtain an estimate of nonverbal intelligence.

Individual Administration of the Reading Test

Only those students who met all criteria for inclusion in this study and had parent permission
received the individual reading test (n=30 high-risk students and n=30 low-risk students). All students were individually administered the Gates-MacGinitie Comprehension section (Level 5/6, Form K). Each student met with the examiner for approximately 1 to 1 1/2 hours, while the reading session was tape recorded for later analysis.

During the individual test administration, the student was instructed to read each passage aloud to the examiner. After reading the passage aloud, the student was instructed to read the question aloud, and to silently choose his or her answer and mark it directly in the test booklet with the pencil provided. Immediately following the selection of the answer for that passage, the student was asked to provide a verbal rationale for the selection of that answer choice. If the student did not provide a clear or understandable verbal explanation, the examiner provided a neutral probe to the student, for example, "Tell me why you think that answer is right," "Explain what you mean," "Did you use any other ways to get that answer?" and so forth. Only neutral prompts were used to preclude the possibility of directed probes leading to student perception of desired responses. During oral reading,
the examiner anecdotally noted on the Individual Observation Sheet any nonverbal or overt behaviors exhibited by the student that indicated test-taking methods or problems that occurred. All procedures specified above were followed for each item regardless of the correctness of the response.

The following is the script that was used to explain the testing procedures to each student:

In school you take many tests like this in a group, that have multiple-choice answers. I am having you take this test with me so I can see how and why students pick certain answers to these reading questions.

First, we will do two samples so you understand what I want you to do. Read the whole passage out loud, then read the question out loud, then choose your answer silently and mark the test booklet.

Then, I am going to ask you why you chose that answer, and I want you to give me the reason or reasons. If I am not sure what you mean, I may ask you more questions. I am not allowed to tell you if the answer you picked is right or wrong, and I am not allowed to help you on the test, since that is in the rules. But please do not be worried or nervous because this will not affect your school grades. Just do the best you can, since I want to see how and why you answer these questions as you do. I am using this tape recorder so I can remember what you tell me, otherwise I might forget since many other students will be taking this test with me. Do you understand? Do you have any questions before we begin?
Two practice items included in the Gates-MacGinitie comprehension test were then completed to clarify the procedure for the examinee. In addition to the regular test items, the practice items also were tape recorded. At this time, any further explanation of the procedure was provided if needed. Each student was given a 5-minute break after item 24, which indicated that one-half of the test was completed. The break was necessary to maintain the attention of the students, given the somewhat tiring nature of oral reading.

**Pilot test.** A complete trial of individual reading testing procedures was conducted with a fifth-grade student to serve as a pilot for further data collection. This pilot test was conducted to determine the value and usefulness of the categories developed by the experimenter to use in data analysis (see subsequent section) and to determine if additional categories should be added. In addition, the pilot was used to test the comprehensibility of script directions and to test coding procedures.
Data Analysis

Coding of Errors

Based on analysis of the tape-recorded readings and interactions, and on observations during the individual test administration, types of examinees' errors on test items answered incorrectly were categorized and frequency and percentage of occurrence for each type was computed. This error analysis was conducted for incorrect answer selections. Errors were included in the analysis and frequency count only when the error was deemed to have been instrumental in causing the incorrect answer choice. The following is a description of the categorical system developed for this study, described in terms of specific variables to be examined.

Variable A: Word Recognition/Word Identification

Errors were defined as errors in reading individual words in the passage or distractors (i.e., the written possible answer choices) determined during the oral administration of the test or when the student provided a rationale for an answer choice, or during probes. Errors in the use of phonic analysis, structural analysis, or with use of context clues were subsumed under this category. Word identification errors were
determined through observation by the experimenter and confirmed through analysis of the readings and discussions recorded on the audiotapes. The number of word identification errors were tabulated for each reading passage and totaled for the entire comprehension test for each subject. The following behaviors (observed or stated) were counted as word recognition/word identification errors if deemed instrumental in the student's wrong answer choice:

1. Failure to read a word correctly through inappropriate application of phonic analysis strategies
   (Example: word dinosaur appeared in text. Student attempted to "sound out" the word, but pronounced it dinner).

2. Substitution of one word for another that affected meaning
   (Example: word head appeared in text. Student read it as heard).
3. Failure to read a word correctly because of inappropriately synthesizing word parts (i.e., faulty structural analysis) (Example: word *sheepherder* appeared in text. Student read it as *sheephearlder*).

4. Omission of a word through failure to attempt the word or to apply a useful strategy. (Example: Student exhibited a long pause, and finally gave up and skipped the word).

5. Substitution of a nonsense word for the text word (Example: word *combing* appeared in text. Student read it as *clombing*).

Variable B: **Comprehension Errors** were defined as inability to select the correct answer due to a misunderstanding or misinterpretation of the passage, question, or distractors caused by lack of background knowledge or a misapplication of background knowledge. In addition, this category included literal comprehension and inferential comprehension errors due to inappropriate reasoning. Comprehension errors can be described as primarily reader-based or primarily
text-based. Reader-based comprehension errors are due mainly to faulty or inappropriate background knowledge, while text-based comprehension errors are due mainly to inability to extract the author’s explicit or implicit message. These errors were counted each time they occurred and totaled for the entire comprehension test for each subject. The following are behaviors that were categorized as comprehension errors when instrumental in a wrong answer choice.

1. Reader-based errors: Student clearly stated inappropriate information that led to his or her answer choice based on wrong prior knowledge or lack of prior knowledge, either stated by the student or inferred by the experimenter.

Example A: The text discussed how a watchman from times past made noise to scare away thieves. The question asked "What does the watchman do at night?" The student responds, "I do not know anything about watchmen or what they do at night."

Example B: The text mentioned a strong-looking girl among other information. A question asked "What was unusual?" The child chose the wrong answer choice "the strength of the girl". His rationale was "because I
know that no girls are strong" (Adam, pilot test, April 24, 1990).

2. **Text-based errors:** When asked for a rationale, if it was stated by the student or can be confidently inferred, that the student missed the point of the passage or misunderstood what was read.

Example A: The passage described a girl running fast, and stated, "The wind was combing her hair with its fingers."
The question asked, "When she ran, Jenny’s hair was... (correct answer: "blown back") the student incorrectly chose "shining." When asked for a rationale, the student states, "I did not understand what it meant by "the wind was combing her hair with its fingers."

Example B (literal comprehension): The passage described a couple walking down a path, a girl and a child, whom she was leading by the hand. When asked what the child was doing, the student incorrectly chose "leading." When queried, he stated, "It says the child is leading by the hand" (Adam, pilot test, April 24, 1990).
Example C (inferential comprehension): The passage described a girl running fast, and stated, "The wind was combing her hair with its fingers." The question asked, "When she ran, Jenny's hair was..., the student incorrectly chose "caught in her fingers." When queried he said, "she was combing hair with her fingers" (Adam, pilot test, April 24, 1990).

Variable C: **Meaning Vocabulary Errors** were defined as a student's inability to determine the meaning of a word, although the student experienced no difficulty recognizing or decoding that word. During any part of the procedure, if it appeared the student had difficulty with a word, but recognized or decoded it, the examiner asked if the student knew the word meaning. The following behaviors were classified as meaning vocabulary errors: The student recognized or identified the word, but,

1. the student spontaneously reported not knowing what a word means

2. the student was unable to define a specific word when asked by the experimenter
Coding of Strategic Factors

Strategies were assessed in a similar manner as error types, with the exception that strategies were largely observed and not reported. Therefore, the coding of strategies was conducted through careful observation by the examiner of a student's test-taking behaviors. In addition, the taped recordings did not include information on strategy use, since much of the strategy assessment was purely observational data during the individual test administration.

For strategy use to have been judged to be in evidence, the student had to either verbally report a strategy or be observed to engage in the specified strategic categories listed below. During testing, the behaviors were recorded on the Individual Observation Sheet.

Variable D: **Strategic Factors** considered in the analysis were test-taking strategies. The following are specifications of behaviors and examples. Each component of the strategies are described below, along with a final compilation of the actual strategies used.
1. The student read all answer choices before selecting an answer. This was determined through verbal report by the student, observation by the examiner, finger pointing to the choices, pointing with a pencil, reading the choices quietly to self, or spontaneously reading aloud the answer choices.

2. The student referred to the passage to search for or verify an answer. This was determined through verbal report by the student, observation by the examiner, finger pointing to the passage, pointing with a pencil, or rereading a portion of the passage aloud or quietly to self.

3. The student used elimination of other options to arrive at a final answer. This was determined through verbal report of the student, observation by the examiner, finger pointing, pointing with a pencil, or verbalizations to self that indicate a process of elimination.

(Example: When a student selected a choice, he or she said, "I know it was not C or D, so I chose B.")

4. The student matched words in the answer choices to words in the passage with no reference to their meaning. This was determined through verbal
report of the student or observation by the examiner during any process of the testing.

(Example: The passage mentioned an Indian Chief Joseph, who had dreaded the moment of telling his people to leave the area. The question asked: "The chiefs thought that Joseph was... (correct answer: "right") the student choose "dreaded." When queried, he responded, "Because the word dreaded is up here in the passage."

5. The student guessed an answer. This was determined through verbal report of the student at any time during the testing.

(Example: When queried as to the rationale for an answer choice, the student responded, "I had no idea what the answer was, so I just took a guess.")

6. Immediate answer selection. This was determined to occur if the student immediately chose an answer after the reading selection as if he or she "knew" the answer. The student did not look back to the passage, read all answer choices, nor perform any other strategy in an attempt to "figure out" the answer as reported by the student or observed by the examiner.
Subsequent to collection of all data, it was determined that students did not use the above categories of strategies in isolation. Often combinations of 1, 2, or 3 strategies were used to select an answer. This resulted in collapsing all the data into seven mutually exclusive categories of strategies which were then used in the data analysis. The following are the resultant seven strategies (also reproduced in Table 2):

Strategy 1:  Read all answer choices  
Strategy 2:  Looked back to the passage to find the answer  
Strategy 3:  Read all answer choices and looked back to the passage  
Strategy 4:  Read all answer choices, but then guessed or looked back and matched words  
Strategy 5:  Read all answer choices and used a process of elimination to decide answer  
Strategy 6:  Immediate answer selection—no other strategy used  
Strategy 7:  Looked back to passage and matched words, or guessed or matched words alone

Comparison of Answer Corrections during Individual Testing

A factor that may impact a student’s test performance is the number of answer changes made by the student during individual testing as a result of the examiner’s neutral probes. The examiner’s probe may induce the child to reconsider his or her initial
answer choice, and as a result, the student may make an answer change. In addition, students may differ in type of answer changes, for example from correct answers to incorrect, or incorrect to incorrect or incorrect to correct. Differences in frequency of answer changes during individual testing were compared for good readers and poor readers.

Specifically, answer change was defined as: after an initial choice has been selected, the appropriate circle blackened in the test booklet by the student, and the examiner has probed for the reason that answer choice was selected, the student decided that he or she had chosen the wrong answer, and selected a different answer choice. Only the rationale for the final answer choice was used in the analysis of errors. Each answer change was noted by the examiner on the Individual Observation Sheet and Strategy Checklist (Appendix B).

Reliability of Data

Reliability of the experimenter’s error categorization, as well as adherence to stated procedures was assessed by an independent rater. To assess the reliability of the coding procedures, the rater independently coded student errors (using the
Error Coding Sheet--see Appendix C) by reviewing the audiotapes of the sessions for the 20% of the randomly selected protocols. Before coding, the independent rater was trained by the experimenter. The experimenter explained the procedures to the rater and he was provided with written and oral descriptions of each category, plus examples to accompany each error type. The rater was able to demonstrate competence in the rating procedure by completing 6 training items with 100% accuracy before his coding was undertaken. The rater was selected so that he also possessed a similar background in training and education level as the experimenter.

In addition, to assess the integrity of the experimental procedures, while the taped assessment sessions were reviewed, the rater also noted on a checklist the following items: (a) experimenter adhered to the script, (b) the passage was read orally, (c) the question was read orally, (d) answer choices were read silently, (e) student explained answer choice, (f) student defined a word--if necessary, (g) examiner gave a break after item #24, (h) only neutral prompts were provided by the examiner, (i) the experimenter provided no assistance to the student, and
(j) no correct answers were given to the student during the administration (see Appendix D for checklist rater will employ in this latter task).

Statistical Analysis

Descriptive statistics were calculated and presented for each error category, overall strategy use, strategy use for correct answers, and strategy use for incorrect answers for average readers (low-risk) and poor readers (high-risk).

To further examine group differences among the profiles of error type and strategy use, several Split-Plot Repeated Measures Analyses of Variance were performed. These analyses yielded information on the comparisons of the profiles of the two groups. However, steps were taken to insure that the data were appropriate to perform the analyses of variance and accurately interpret the results. The nature of the data (proportional), necessitated using an arcsin transformation (Kenny, 1987) by computer to normalize and convert the data to standardized units. According to Kenny (1987) this type of two-stretch transformation is appropriate when working with proportional data.
(which has a clear upper bound of 1.0 and lower bound of 0.0), since it stretches the distributions' two tails which results in normalization. The transformed data were then used to perform the ANOVAS.

It is suggested by Kennedy and Bush (1985) and Kenny (1987) that often when samples are small, it is difficult to assess the degree to which the distributions approach normality. In addition, often the homogeneity of variance assumption needed to perform the analysis of variance is violated. According to Kennedy and Bush (1985) if the assumption of homogeneity of variance is violated, the resultant effect is that of a positively biased $F$ statistic, which indicates increased probability of type I errors. Similarly, when working with smaller samples, the assumption of normally distributed data is often violated. This violation also leads to a positively biased $F$, which increases the chance of committing a type I error.

In regard to the violation of the normality assumption, Kennedy and Bush (1985) cited several studies which tested the effect of violation of this assumption. Based on results of these studies, they concluded, "it is reasonable to conclude that
departures from population normality have little effect, in practice, relative to spuriously inflating the probability of committing type I errors" p. 146. They also stated, "Thus, the analysis of variance appears to be extremely robust with respect to violations of the normality assumption, a finding most gratifying to behavioral researchers who sometimes find that population non-normality is the rule rather than the exception" p. 146.

In response to the violation of the homogeneity of variance assumption, and the violation of the normality assumption, Kennedy and Bush (1985) stated, "researchers should be sensitive to the nature of the problem and should initiate steps to compensate for positive bias if there is the slightest possibility of advancing an erroneous inference" (p. 383). The method suggested by Kennedy and Bush is that of applying a conservative $F$ test. Kennedy and Bush stated,

Many methods have been proposed to correct the $F$ statistic for positive bias. Of those, the Geisser-Greenhouse conservative $F$ test will be satisfactory for most work. This test reduces the degrees of freedom in such a way that the $F$ is adjusted conservatively. The alpha rate is protected at the nominal level even if data exhibit the worst possible violations of the homogeneity assumptions. (p. 383)
Error profiles of good and poor readers were compared using a $2 \times 3$ (group by error category) split-plot repeated measures ANOVA. The repeated measure analyses were tested using the Geisser and Greenhouse (1958) adjusted degrees of freedom test which provides a conservative measure of alpha level. Specific differences between groups on each variable were assessed by performing three selected t-tests using the Bonferroni correction technique (Games, 1971) to protect alpha level.

Similarly, the strategy-use profiles of good and poor readers were compared first for overall strategy use by using a $2 \times 7$ repeated measures ANOVA. In addition, to measure group differences on correct and incorrect answer choices, two $2 \times 7$ (group by correct/incorrect answer by strategy) split plot repeated measures ANOVAS were performed. All the analyses were tested using the Geisser-Greenhouse adjusted degrees of freedom test, which is conservative. Selected post-hoc pairwise comparisons were performed for those ANOVAS that showed significant main effects. The Scheffe post-hoc test was used to determine specific differences between groups on each variable when correct answers were obtained.
Summary

Sixty sixth-grade students were the participants in this study, 30 high-risk and 30 low-risk students. Subjects were selected based on several factors: group-administered reading test scores, participation in the free lunch program (for high-risk subjects), and parent permission. The testing was performed in a large midwestern city and two nearby suburbs, and was conducted by an experimenter experienced in assessment.

The main source of data was the individually-administered reading test (Gates-MacGinitie Reading Tests, Level 5/6, Form K) for which students were required to read the test aloud and provide a rationale for each of their answer choices. The type of error deemed instrumental in causing a wrong answer choice was recorded according to the error categorization scheme described in this chapter. In addition to errors, strategy use also was assessed. Students were evaluated on the basis of strategies used according to the categories described in this chapter.

Reliability of error coding and adherence to stated procedures, was assessed through use of a trained, qualified independent rater.
Profile differences of good readers (low-risk) and poor readers (high-risk) on error types, strategy use, and answer corrections were examined. Descriptive statistics were calculated for both groups, and several split-plot, repeated-measures analyses of variance were performed to test stated research questions (see chapter II). Conservative F tests were employed to protect alpha level given possible violations of the normality and homogeneity of variance assumptions. Selected post-hoc tests were used to compare specific variables in question.
CHAPTER IV

RESULTS

This chapter provides the results of comparisons between poor readers and average readers on the variables of error types and strategy use. The chapter is divided into four sections: (a) interrater reliability (b) errors between groups (c) strategy use between groups, and (d) answer changes.

Interrater Reliability

To assess the reliability of the data in error categories as rated by the experimenter, an independent rater randomly selected and coded the student errors by listening to 20\% of the audiotaped readings. The same error coding sheet employed by the experimenter was used. (See Appendix C). The overall interrater reliability coefficient was .90 between the examiner's categorizations and the independent rater's categorizations.
Specifically, the inter-rater reliability coefficients for comprehension, word identification, and meaning vocabulary error categorizations were .98, .98, and .75 respectively. There was more disagreement between the examiner and the independent rater on the categorization of meaning vocabulary errors.

To assess the integrity of the testing procedures, a checklist of correct procedures was also employed for each tape reviewed by the independent rater. (See Appendix D). The rater found that overall the correct procedures were followed 98% of the time for the 20% sample that was analyzed. Specifically, 100% adherence to procedures for all categories with one exception was found. The category "student defined a word when needed" resulted in agreement 82% of the time between the examiner and the independent rater. It was felt by the rater that the examiner should have queried students on the meanings of words more frequently in certain cases. This disagreement also impacted the assessment of meaning vocabulary errors in this study since the student's definition of a word was directly linked to the scoring of a meaning vocabulary error. This indicates that the incidence of meaning vocabulary errors may be underestimated in this study.
Error Types Between Groups: Poor Readers versus Average Readers

Research Question 1 addressed the relationship between error types and reading ability. Specifically, within this research question, hypotheses 1-3 stated that there would be significant differences between groups when word identification errors, comprehension errors, and meaning vocabulary errors were compared. Since poor readers made more erroneous responses, the total proportion of errors of each type was calculated. See Table 1.
Table 1

Frequencies, Means, Standard Deviations, and Percentages of Word Identification (WI), Meaning Vocabulary (MV), and Comprehension (COMP) Errors by Group

<table>
<thead>
<tr>
<th></th>
<th>Average Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq. Mean SD %</td>
<td>Freq. Mean SD %</td>
</tr>
<tr>
<td>COMP</td>
<td>311  10.37  2.82  80.99</td>
<td>477  15.90  5.83  63.68</td>
</tr>
<tr>
<td>MV</td>
<td>46   1.53   1.41  11.98</td>
<td>29   0.97   .93   3.87</td>
</tr>
<tr>
<td>WI</td>
<td>27   .90    1.06  7.03</td>
<td>243  8.10   5.69  32.45</td>
</tr>
<tr>
<td>Tot</td>
<td>384</td>
<td>749</td>
</tr>
</tbody>
</table>

Note. Non-transformed data.

The frequencies and percentages in this table show that poor readers demonstrated more word identification errors and obtained a higher proportion of that error type than average readers. However, average readers made more meaning vocabulary errors and demonstrated a higher percentage of this type of error than poor readers. Comprehension errors were the most frequent error type made by both groups, however, average readers made
proportionally more comprehension errors than poor readers.

To determine if the differences in the profiles of the two groups stated above were significant, a split-plot, repeated measures analysis of variance was performed. As stated in Chapter III, the existing data were converted to percentages of error type used by each individual, then an arcsin transformation performed by computer to convert the proportional data to standard, normalized units. This allowed direct comparisons between the groups despite the disparity in actual frequency of error type.

The results of the analysis of variance indicated there were significant group differences on the error type variables, $F(1, 58), = 10.02, p<.01$. In addition, there was also a significant group by error type interaction, indicating the profiles of group performance differed significantly, $F(1.25, 72.49), = 23.88, p<.001$. Figure 1 represents the profiles of both groups.

Three post-hoc t-tests using the Bonferroni correction were performed to determine specific differences between each error type variable between groups and to address Hypotheses 1-3. These results indicated significant differences between groups existed
on all variables. Using the transformed data, and holding the experimentwise error rate to \( p < .05 \), \( (t-\text{crit}=2.00, \ df=58) \), Comprehension Error means differed significantly, (Group 1-Average Readers Mean=1.04; Group 2-Poor Readers Mean=.72; minimum required for significant difference = .15, \( p < .05 \)). Meaning Vocabulary Errors differed significantly, (Mean Group 1 = .11; Mean Group 2 = .04; Minimum required for significant difference = .04, \( p < .05 \)). Word Identification Errors differed significantly, (Mean Group 1 = .07; Mean Group 2 = .34; Minimum required for significant difference = .08; \( p < .05 \)).

These results indicate there were group differences on each of the variables and the overall group profiles also differed significantly. (See Figure 1).

Figure 2 presents the profiles of the two groups superimposed on one graphic for ease of comparison.
Error Type

<table>
<thead>
<tr>
<th></th>
<th>Comprehension</th>
<th>Meaning</th>
<th>Word Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Readers</td>
<td>.8</td>
<td>.6</td>
<td>.4</td>
</tr>
</tbody>
</table>

FIGURE 1. A Comparison of Profiles (Mean Transformed Proportions) of Error Types by Group
Figure 2. Mean (Transformed Proportions) Scores of Average and Poor Readers by Error Type
Strategy Use Between Groups: Poor Readers versus Average Readers

**Overall Strategy Use**

In Research Question 2, it was asked if poor readers differ from good readers in use of strategies. See Table 2 for a description of each strategy type.

Table 3 presents results of strategy use by both poor readers and average readers, regardless of correctness of the response selection. Examination of frequencies and percentages in this table indicate that both groups often used strategies 1, 3, and 6. However, when examining strategy use, average readers used strategy 1 most often, followed by strategy 3, then strategy 6. Poor readers also used strategy 1 most often, followed by strategy 6, then strategy 3, then strategy 4. Strategies 2, 5, and 7 were used relatively infrequently by either group.

To determine if differences between groups were significant, a split-plot, repeated measures analysis of variance was performed.

The results of the analysis of variance indicated there were significant group differences on the strategy use variables, $F(1, 58) = 5.42, p < .05$. In addition, there was also a
# Table 2

**A Description of Strategies**

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy 1</strong></td>
<td>Read all answer choices</td>
</tr>
<tr>
<td><strong>Strategy 2</strong></td>
<td>Looked back to passage to find answer</td>
</tr>
<tr>
<td><strong>Strategy 3</strong></td>
<td>Read all answer choices and looked back to passage</td>
</tr>
<tr>
<td><strong>Strategy 4</strong></td>
<td>Read all answer choices, but then guessed or looked back and matched words.</td>
</tr>
<tr>
<td><strong>Strategy 5</strong></td>
<td>Read all answer choices and used a process of elimination to decide answer</td>
</tr>
<tr>
<td><strong>Strategy 6</strong></td>
<td>Immediate answer selection—no other strategy used</td>
</tr>
<tr>
<td><strong>Strategy 7</strong></td>
<td>Looked back to passage and matched words, or guessed, or matched words</td>
</tr>
</tbody>
</table>
Table 3
TOTAL FREQUENCY AND PERCENTAGE OF STRATEGY USE BY GROUP

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>AVERAGE READERS</th>
<th>POOR READERS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FREQ.</td>
<td>MEAN</td>
</tr>
<tr>
<td>1</td>
<td>653</td>
<td>21.77</td>
</tr>
<tr>
<td>2</td>
<td>49</td>
<td>1.63</td>
</tr>
<tr>
<td>3</td>
<td>240</td>
<td>8.00</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>1.93</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>.90</td>
</tr>
<tr>
<td>6</td>
<td>388</td>
<td>12.93</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>.23</td>
</tr>
<tr>
<td>TOT</td>
<td>1440</td>
<td></td>
</tr>
</tbody>
</table>

Note. Non-transformed data.
significant group by strategy type interaction, indicating the profiles of group performance differed significantly, \( F(1.79, 103.70), = 12.20, p < .001 \). Figure 3 represents the profiles of both groups. A post-hoc analysis was performed using the Scheffe method of multiple comparisons. Results indicated that significant differences existed between strategies 1, 2, 4, and 6. (See Figures 3 and 4).

**Strategy Use and Correct Answer Selection**

Research Question 3 addressed the effectiveness of strategy use, and it was asked in what way poor readers and average readers differ in strategy use when students obtained a correct answer? Table 4 displays data regarding the relationship between strategy use and correct answer responding before the transformation was performed. Percentages listed in this table indicate that strategies 1, 3, and 6 appeared to be used most frequently, relative to the other strategies, for both groups when correct answers were selected.
FIGURE 3. A Comparison of Profiles (Mean Transformed Proportions) of Overall Strategy Use by Group
Figure 4. Mean Transformed Proportions of Average and Poor Readers for Overall Strategy Use
Table 4

STRATEGY USE WHEN STUDENTS ANSWERED QUESTIONS CORRECTLY

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>AVERAGE READERS</th>
<th>POOR READERS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FREQ.</td>
<td>MEAN</td>
</tr>
<tr>
<td>1</td>
<td>466</td>
<td>15.53</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>1.30</td>
</tr>
<tr>
<td>3</td>
<td>181</td>
<td>6.03</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>.17</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>.70</td>
</tr>
<tr>
<td>6</td>
<td>342</td>
<td>11.40</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>.07</td>
</tr>
<tr>
<td>TOT</td>
<td>1056</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Non-transformed data.
To determine if there were significant group differences, a repeated measures, split plot analysis of variance was performed. Again, as for the error types, due to the disparity in frequencies between groups, the data were converted to percentage of use of each strategy for each student, then an arcsin transformation performed to convert the data to normalized units.

The results of the analysis of variance indicated there was a significant main effect between groups on the variables of strategy type for correct responding, \( F(1, 58), = 27.47, p<.01 \). In addition, there was a significant group by strategy type interaction, indicating the profiles of group performance differed significantly, \( F(4.17, 241.67), = 3.62, p<.01 \).

Figure 5 represents the profiles of both groups.

Since a significant main effect between groups was obtained, a Scheffe post-hoc test was performed to make pairwise comparisons between groups on each variable. These results indicated that significant differences between groups existed on: Strategy 1, Strategy 2, and Strategy 3, and no significant differences existed on Strategies 4-7. (See Figure 6).
FIGURE 5. A Comparison of Profiles (Transformed Mean Proportions) of Strategy Use for Correct Answers by Group
Figure 6. Mean Scores of Average and Poor Readers for Strategy Use When Students Selected Correct Answers
Strategy Use and Incorrect Answer Selection

Research Question 4 addressed ineffective strategy use, and it was asked what differences exist between groups when students selected an incorrect answer. Frequencies and percentages in Table 5 indicate that strategies 1 and 4 were used most frequently by poor readers, and strategy 1 was used most frequently by average readers when incorrect answer choices were selected.

To determine if significant differences existed between groups, a repeated measures, split-plot analysis of variance was performed. The results of the analysis of variance indicated there were no significant main effects between groups on the variables of strategy type for incorrect answer selection, $F(1,58), = 3.84, p>.05$. However, there was a significant group by strategy type interaction, indicating the profiles of group performance differed significantly, $F(4.34, 251.52), = 4.34, p<.05$. Figure 7 represents the profiles of both groups.
Table 5

STRATEGY USE WHEN STUDENTS ANSWERED QUESTIONS INCORRECTLY

<table>
<thead>
<tr>
<th>STRATEGY</th>
<th>AVERAGE READERS</th>
<th>POOR READERS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FREQ.</td>
<td>MEAN</td>
</tr>
<tr>
<td>1</td>
<td>187</td>
<td>6.23</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>.33</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
<td>1.97</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>1.77</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>.20</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
<td>1.53</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>.77</td>
</tr>
<tr>
<td>TOT</td>
<td>384</td>
<td>747</td>
</tr>
</tbody>
</table>

Note. Non-transformed data.
FIGURE 7. A Comparison of Profiles (Transformed Mean Proportions) of Strategy Use by Group When Incorrect Answers Were Chosen
Because there was no overall group differences, this demonstrates that both good and poor readers used the same strategies when incorrect answers were selected. Both groups often used strategy 4 when an incorrect answer choice was selected. However, the interaction pattern determined that the profiles of the groups differed on strategies 6 and 7. Therefore, poor readers and good readers differentially used strategies 6 and 7 when wrong answers were selected. Average readers used strategy 7 more often than poor readers, and poor readers used strategy 6 more often than average readers when incorrect answers were selected. This shows that average readers continued to use the strategy of looking back to the passage, although when coupled with matching words or guessing (strategy 7), they were unsuccessful, and often obtained incorrect answers. When poor readers used immediate answer selection (strategy 6), they more often obtained incorrect answers than average readers. (See Figure 8).
Figure 8. Mean Transformed Proportions for Average and Poor Readers for Incorrect Answer Responding
Answer Changes Between Groups

In Research Question 5, and the four subsequent hypotheses statements, it was asked if there was a difference between the poor readers and the average readers when the frequency of answer changes was compared. To investigate differences between groups when students initially selected an answer, then changed the answer, a comparison was made between the number of answer changes for each group. Total frequency of answer changes did not significantly differ between poor readers ($M = 0.90; \ SD = 1.32$) and average readers ($M = 0.73; \ SD = 1.11$), $t(58) = 0.53, p<.5993$.

Furthermore, when average readers and poor readers were compared on three specific types of answer changes ((a) incorrect to correct, (b) incorrect to incorrect, and (c) correct to incorrect), they were found to differ only when answers were changed from correct to incorrect. Poor readers ($M = .27; \ SD = .45$) changed answers from correct to incorrect significantly more often than average readers ($M = .06; \ SD = .25$), $t(58) = 2.12, p<.05$. No significant differences were found between average and poor readers when answers were changed from incorrect to correct ($t(58) = .31, p<.76$), or when answers were changed from incorrect to incorrect ($t(58) = .66, p<.51$).
Summary

The results of this study showed significant differences between average and poor readers on several reading and test-taking variables. Specifically, average readers and poor readers exhibited significantly different patterns of error types. Average readers exhibited significantly more comprehension errors, and less word identification errors than poor readers. Additionally, it was found that meaning vocabulary errors were more frequently the cause of a wrong answer selection for average readers than for poor readers.

A difference between average and poor readers was also found for overall strategy use. While poor readers tended to read all the answer choices (strategy 1), they often coupled that strategy with either guessing (strategy 4) an answer or matching words (strategy 4) to a greater degree than average readers. Conversely, average readers tended to use immediate answer selection (strategy 6) and looking back to the passage (strategy 2) more than poor readers.

When strategy use between average and poor readers was compared using correct and incorrect answer selections, significant differences occurred. When students obtained correct answers, average readers were
found to use strategies 1, 2, and 3 more than poor readers. This indicates that average readers make better use of reading all the answer choices (strategy 1), looking back to the passage, (strategy 2) and using a combination (strategy 3) of those two strategies when correct answers were selected. Conversely, when students selected incorrect answers, both groups often used strategy 4, that is reading all the answer choices, then guessing or matching words. Average readers used strategy 7 (looking back to the passage, then guessing or matching words) significantly more than poor readers. Poor readers used strategy 6 (immediate answer selection) more than average readers when incorrect answers were obtained.

The propensity to alter initial answer selection was also compared between groups. It was found that good and poor readers did not differ on their overall frequency of answer changes. Additionally, no differences on changes made from incorrect to correct, or incorrect to incorrect were found. However, significantly more poor readers changed their initial answer selection from correct to incorrect than average readers.
CHAPTER V
DISCUSSION

In the present study, error types and test-taking strategies of good readers and poor readers were compared. Several data analyses were used to determine group differences on the above variables, which were reported in Chapter IV. In this chapter, findings for all research questions are discussed. This chapter is divided into eight sections, (a) error types between groups, (b) overall strategy use, (c) strategy use and correct answers, (d) strategy use and incorrect answers, (e) strategy use and cognitive processing (f) answer changes, (g) limitations of the study, (h) directions for future research, and (i) summary.

Error Types Between Groups

When poor readers and average readers were compared on relative performance of comprehension errors, word identification errors, and meaning vocabulary errors on the reading comprehension test, differing patterns
resulted. It was apparent that average readers and poor readers both make mistakes of all types, however in differing frequencies (proportions). When the disparity in number of incorrect answers was controlled, the proportion of errors due to word identification problems was significantly higher for poor readers than for average readers. This indicates that, for poor readers, word identification difficulties are disproportionately interfering with answering comprehension questions accurately. This finding is consistent with contentions by Stanovich (1991), and others who support interactive reading theories, that word recognition and word identification skill is a fundamental part of reading comprehension, and also their contentions that it is an effective discriminator between average and poor readers. Poor word recognition skill (which was demonstrated more often by poor readers than average readers) significantly and negatively impacted students’ performance on the comprehension test. This raises questions as to the use of comprehension tests that are beyond the reading level of the student, and raises questions about the amount of actual comprehension measured by comprehension tests for the poor reader group. Much support has been found for out-of-level testing if one is to measure a student’s
comprehension skill independent of word identification skill (Fisher, 1962).

In addition, average readers made significantly more errors due to faulty knowledge of word meanings (meaning vocabulary) than poor readers. This is not surprising, since poor readers had greater difficulty with word identification. Because poor readers had difficulty decoding words, they had no opportunity to demonstrate whether or not they possessed knowledge of a word's meaning. For example, the coding scheme developed in this study was designed in such a manner so that to code a meaning vocabulary error, the student must first be able to decode the word. In this case, meaning vocabulary errors were hierarchically more advanced than word identification errors. However, meaning vocabulary error categorizations and examiner probes for word meanings were not as reliable in this study as the other categories. Therefore, it was difficult to know the extent to which poor readers have meaning vocabulary difficulty.

These results indicate that meaning vocabulary errors negatively affect test performance, although more frequently for average readers than poor readers. This information is congruent with the results of research by Drum, Calfee, and Cook (1981) and Anderson and Freebody
indicating performance on a standardized comprehension test decreases when the number of unfamiliar words increases. Although not specifically addressed in this study, it is likely that word identification difficulties also negatively impact vocabulary portions of standardized tests that are specifically designed to assess knowledge of word meanings.

In the area of comprehension, it was found that the proportion of errors due to comprehension problems was greater for average readers than poor readers. This is not surprising given the disproportionate negative influence of word recognition errors demonstrated by poor readers. Word recognition errors prohibit poor readers from making errors due to comprehension failure, which appears to be a more advanced level of understanding. That is, students cannot even approach the task of comprehension without recognition of the words on the printed page. This is true in instructional reading situations and on reading tests. It also was true in this study. Furthermore, with the coding scheme developed in this study, a student was categorized as having made one of the three mutually exclusive types of errors. If poor readers made a significant number of word identification errors, the number of comprehension errors would be
negatively affected since they were only credited with making one type of error per question. Therefore, the comprehension test seems to be a more accurate reflection of reading comprehension with average readers than poor readers.

Overall these results on the differential pattern of error types for average and poor readers support recent findings in reading research on the importance of fluent, context-free word recognition (Stanovich, 1991), as well as well-developed word identification strategies (Adams, 1989). The field of reading has vacillated from a focus on word identification as a first step in bottom-up processing to a top-down philosophy minimizing an emphasis on explicit teaching of word identification.

More recently, the focus has been on the interaction between word identification and comprehension in information processing models. For example, several studies investigating the relationship between successive and simultaneous cognitive processing and reading difficulties have been cited by Das, Kirby, and Jarman (1979). These studies have indicated that successive processing is more related to reading decoding, and simultaneous processing is more related to reading
comprehension. However, these studies did not measure the planning component of cognitive processing.

The results of the present study indicate that word identification is certainly an integral part of successful reading comprehension, and it is certainly a discriminating factor between average and poor readers. However, other factors do impact reading test performance, which will now be discussed.

Strategy Use Between Groups

Overall Strategy Use

When groups were compared on overall use of strategies, without regard to correct or incorrect answer choices, significant differences in the group profiles of average and poor readers were found. The results indicated that use of strategies 1, 2, 4, and 6 were significantly different, and no differences were found in overall use of strategies 3, 5, and 7. Strategies 1 and 4 were used more frequently by poor readers (although strategy 1 was also used frequently by average readers). For example, they often read all the answer choices before responding, and also combined reading all the answer choices with either guessing or matching words in the answer choice to words in the passage. This is contrary
to the finding by Scruggs, Bennion and Lifson (1985) who found that students often did not read all the answer choices before responding. In this study, however, poor readers engaged in this behavior more than average readers. However, for poor readers, reading all the answer choices seemed to be coupled with the detrimental strategies of guessing or matching words.

Average readers were found to engage in the use of strategies 2 and 6 more often than poor readers when correctness of responses was not analyzed. Although average readers engaged in the use of strategy 2 more often than poor readers, the actual use of that strategy was relatively infrequent. This indicates that average readers more often looked back to the passage to search for or verify an answer, and also engaged in more immediate answer selection, (without use of any other strategies) than poor readers. This result is consistent with research by Scruggs et al. (1985) who found that a number of children did not look back to the passage to search for an answer, and several children engaged in immediate answer selection. This also indicates that better readers may actually need to use fewer strategies to answer comprehension questions, indicating more confidence in their initial answer choices.
Although poor readers used strategy 1 more than average readers, and average readers used strategy 2 more than poor readers, strategy 1 (reading all the answer choices) was actually popular among both groups, and strategy 2 (looking back to the passage-no other strategy used) was not popular for either group. This is consistent with one result reported in a study by Farr et al. (1990) who found that the most popular question-answering strategy was that of reading all the answer choices, then eliminating wrong answers (although little support for elimination of answer choices was found in the present study). In addition, Farr et al. found that looking back to the passage to locate an answer was a popular question-answering strategy for his sample. They concluded that the test questions themselves often guide test-taking behaviors. It was observed in the Farr et al. study that the students often alternated back and forth between the question and the passage, and therefore reading comprehension tests are a special type of reading task. This assertion was somewhat supported by findings in this study, since students engaged in active question-answering strategies.
Strategy Use and Correct Answer Selection: Between Groups

When strategy use between groups was compared on data consisting only of strategies used to obtain correct answers, average readers demonstrated use of significantly more strategies than poor readers. In addition, strategies 1, 2, and 3 were used significantly more by average readers than poor readers when correct answers were selected. There were no significant differences found between the groups for strategies 4, 5, 6, or 7 when correct answers were obtained. This indicates that average readers more successfully used the strategies of: (a) reading all the answer choices, (b) searching in the passage for an answer, and (c) performing a combination of reading all the answer choices and searching in the passage for the answer. Although the poor readers used all the same strategies, they did so less often and less effectively. In addition, strategy 6, although not different between groups, was generally effective when correct answers were obtained.

Strategies 1, 2, and 3 are notably positive strategies, indicating that not only do average readers use the strategies more often, they also used positive
strategies over negative strategies when correct answers were obtained. In reference to reading all the answer choices, Scruggs, Benion and Lifson (1985) cited failure to read all the answer choices in a multiple-choice test as detrimental to reading test performance. In addition, according to Alexander, Hare, and Garner (1984), text lookbacks are beneficial for obtaining correct answers, and are therefore, positive. Garner (1984) found that when students were presented with text and questions together, lookbacks were more beneficial than when text and questions were presented separately.

The results stated above are consistent with those of Kavale and Schreiner (1979) who found better readers to have greater frequency of strategy use. As in the present study, Kletzien (1991) also found that the good and poor comprehenders used the same strategies, but good readers used more variety and controlled them more in the face of difficult passages. This supports the hypothesis that group differences exist in the regulation of strategy use, not in the knowledge of the strategies themselves. In this study, it was apparent that poor readers were as capable of using the same strategies to obtain correct answers as average readers, however, average readers applied them more consistently and effectively. This
finding supports the hypothesis that differences in the regulation of strategies contributes to group differences in reading ability.

**Strategy Use and Incorrect Answers**

No significant overall group differences existed on the strategy type variables when students selected incorrect answers. Both groups often used strategy 4 when an incorrect answer was obtained, which was **reading all the answer choices but then guessing or looking back and matching words in the passage to the words in the answer choices**. Obviously, for both average and poor readers, matching words in the passage with the answer choice is a negative strategy. This strategy has been referred to as "search and destroy" by Pearson (1979). In addition, guessing was also largely negative. While reading all the answer choices seemed to be positive for average readers, it also seems to be a necessary but not sufficient condition for obtaining correct answer choices. If one were to increase her chances of avoiding incorrect responding, it would be better to read all answer choices, look back to the passage to check, then finally, select an answer.
When incorrect answers were selected, interaction patterns showed that average readers and poor readers performed differently on the use of strategies 6 and 7. Strategy 7 was looking back to the passage and matching words, or guessing, or matching words alone (this strategy did not include reading all answer choices). Average readers used strategy 7 more than poor readers. This indicates that average readers continued to employ the strategy of looking back, but the combination of looking back and reading all answer choices was not used when answer selections were incorrect. When they resorted to matching words or guessing, this also reduced their chances of obtaining a correct answer. Strategy 6 was immediate answer selection. This was a successful strategy for correct answers (especially for good readers), however the poor readers obtained a significantly greater amount of incorrect answers than average readers when this strategy was used.

In addition, although not formally measured in this study, poor readers testing times were markedly longer than average readers. This is not surprising with the finding of laborious searching for answers on the part of the poor reader, which usually ended with matched words or
a guess (strategy 4). It was apparent that poor readers do not efficiently engage in text look-backs.

**Strategy Use and Cognitive Processes**

It was apparent in this study that, although all students were capable of using all strategies, the average readers were more able to plan and direct their strategy use than the poor readers. This finding is consistent with recent research in the area of cognitive processing where planning deficits were found to exist in learning disabled students. Studies by Bardos (1988) Saneda and Serafica (1991), and Flanagan (1992), found that students did not display deficits in coding (simultaneous and successive processing) but had difficulty with the cognitive area of planning. That is, learning disabled students displayed deficits in the knowledge and regulation of cognitive activities and the ability to direct strategy use. These findings are consistent with the results of the present study.

Various remediation strategies for global and specific cognitive and academic skill deficits have been investigated by Cormier, Carlson, and Das (1990) and these interventions continue to be developed and tested.
Answer Changes

In Research Question 5 it was asked if there was a difference between groups in frequency of answer changes. Answer changes may be indicative of self-monitoring behavior, since a student is actively questioning himself or herself as to the correctness of a response. The results indicated that there was no difference between the groups on total number of answer changes. Therefore, both groups were attempting to monitor themselves. However, the degree of successfulness of the self-monitoring was found to differ between groups. Poor readers made changes from a correct answer to an incorrect answer significantly more often than good readers, obviously a detrimental change. There was no difference between groups when changes were made from a wrong answer to another wrong answer, nor from a wrong answer to a correct answer. Therefore, poor readers appear less confident in their initial answer selection when they are, in fact, correct. It was observed by the examiner during testing that certain incorrect answer selections were commonly chosen distractors. These frequently chosen distractors usually included specific words or exact statements that occurred in the text. Since poor readers more often altered their initially correct answer selections to those that were
incorrect, it is likely that the poor readers were more easily led by the attractive incorrect distractor items on the multiple-choice test than are average readers.

**Limitations of this Study**

One limitation of the present research involves the development of the categorical scheme itself. The categories were developed by the experimenter, and are certainly not exhaustive of all specific error types and strategies in a student's repertoire of behaviors. Although generally based on other research in the field, in the development of the categories, the researcher may have omitted a strategy type or developed error categories that were too broad. However, after using the categorical scheme to classify all the data, the experimenter felt that the most relevant and important information was captured within the developed categories. In fact, at the time of development of the categories, a category labeled "other" was included to provide information on a strategy not listed. After the collection of data, the category "other" was re-labeled "immediate answer selection." It was felt that the categorical scheme effectively
categorized the behaviors observed throughout the testing sessions.

A second limitation involves reliability of the strategy coding procedure. Since strategy assessment was completed at the time of the test administration, and was often based on observational (nonverbal) data by the examiner, there was no reliability measure of the experimenter's categorizations of strategy use. However, if reliability of strategy assessment were to be measured, it would entail an additional observer present in the room also observing the student closely. It was felt that this would be an unnecessarily artificial and anxiety-provoking situation for the student because of the necessary close proximity of the second observer. Despite having no reliability check of these variables, the examiner felt confident in the use of the coding procedure. The students often exhibited obvious nonverbal cues which denoted specific strategy use. For example, most students pointed with their fingers, pointed with their pencils, subvocalized strategies, or made obvious head and eye movements to indicate strategy use. This made strategy assessment obvious the great majority of the time.
A third limitation involved the initial selection of students. Initially it was proposed that all students were to be selected by administration of the same measure, the Gates-MacGinitie Reading Comprehension Tests, Third Edition, Level 5/6, Form L (MacGinitie & MacGinitie, 1989). However, in two schools where data from all the poor readers and approximately one-third of the average readers was collected, this was not allowed by school district policy. The researcher was required to use existing test data (comprehension scores only) from the Comprehensive Test of Basic Skills (CTBS; CTB/McGraw Hill, 1977) to select the poor readers and some of the average readers. The disparity in selection instruments may be a disadvantage if the tests are not comparable. However, the CTBS has a multiple-choice reading comprehension format similar to the Gates MacGinitie, the CTBS had been recently administered to the students (previous spring), and standard scores were used to select students from both tests.

A fourth limitation of this study is the degree to which the data collection procedures resembled actual student test-taking errors and strategies performed in school. Often students are more motivated when performing individually than in a group setting. Some of the
strategy use or problem-solving behaviors of students may have been increased as an attempt to perform better in the presence of an examiner. Therefore when individual test administration is compared to group test administration, students may perform some different behaviors. However, an individual-administration methodology was necessary to investigate the research questions addressed in this study.

A fifth limitation of this study is related to sample size. Although 60 subjects (30 average readers and 30 poor readers) would be considered a fairly large sample when this type of individual, in-depth, protocol analysis is involved, one should exercise caution when making generalizations from the results. The findings are valid to the degree that this sample is representative of average and poor readers in general, and the larger the sample size the more likely representativeness will occur. Great care was taken to select each student so that all inclusion criteria were met in this study. However, statistical power is somewhat limited with the sample size used, and the reader of this report should be cognizant of the fact the results are based on 60 students.
A sixth limitation involved the use of the Matrix Analogies Test-Expanded Form (MAT) to assess intellectual ability. The MAT measures only nonverbal (simultaneous processing) ability. Therefore, it was not surprising that all students obtained scores within the average range. Since other components of cognitive processing such as successive processing, planning, and attention were not measured in this study, it is difficult to specifically address their relationship to reading difficulty.

Directions for Future Research

Information gained in this study on strategy use associated with good and poor readers may have important educational implications. Future research might investigate an intervention in which readers are trained in the use of effective test-taking strategies. This might include teaching students strategies such as those implicated by this study in successful performance. That is, if they do not know the answer immediately with a fair amount of certainty, then reading all answer choices and looking back to the passage to verify or search for an answer is effective in obtaining correct answers. This overall strategy is a combination of those used most often by both good and poor readers in the present study when
answer selection was correct. The knowledge of this strategy should be promoted, and consistent use of that strategy by poor readers should be encouraged. The effectiveness of this intervention could then be evaluated.

It was apparent that the intermediate level sixth-grade students in this study used a variety of test-taking strategies, and demonstrated some flexibility of strategy use. Future research with differing age levels, such as primary age and secondary level students, would provide information on developmental trends in amount and type of strategy use. Early intervention at the primary grades would likely be beneficial, since by sixth grade it is apparent that students are still using a number of detrimental test-taking strategies.

Since this study investigated strategy use and error types of average and poor readers, additional information on superior readers might be beneficial to the field. If the performance of superior readers was examined during a multiple-choice comprehension test, an indication of the highly effective strategies used by these readers could be ascertained. These strategies used by superior readers could then be taught to average and poor readers in an attempt to increase their test performance.
It was found in this study that word identification errors were more frequently exhibited by poor readers than average readers, and that this error type had a detrimental effect on comprehension, especially so for the poor reading group. Further research on the effects of word identification errors on comprehension test performance seems to have merit. For example, if an experimenter provided students with the pronunciation and/or meaning of each word they were having difficulty identifying, it seems that a truer measure of comprehension could be attained. Comparing this procedure with that employed in the present study (no use of experimenter prompts) could test the validity of this assertion.

In this study, several average readers engaged in immediate answer selection, with no use of additional strategies. In addition, testing times for poor readers were markedly longer than for average readers. A study investigating the use of superior readers' test-taking strategies on a comprehension test could be conducted to shed light on whether or not superior readers actually engage in test-taking strategies, or if it is the absence of strategies that is a contributing factor to swift, accurate responding.
Since support for interactions between different processes of reading was found in this study, the type of information gathering employed in the present study could be taken a step further and used to evaluate an individual student for specific strengths and weaknesses in reading. In addition, to determine other cognitive processing strengths and weaknesses that may contribute to reading difficulty, this information could be combined with information gathered through additional measures of cognitive processing used for individual student evaluation. This would provide an examiner not only with information on what mistakes students are making during test-taking, but also why some of these difficulties are occurring within a cognitive processing model of intellectual functioning.

Summary

This study investigated the error types and test-taking strategies demonstrated by high-risk (poor readers) and low-risk (average readers) students on a standardized, multiple-choice comprehension test. Thirty students classified as high-risk and thirty low-risk students participated in the study. A comprehensive error and strategy coding procedure was developed by the
experimenter prior to individual testing. Students were then individually tested using the Gates-MacGinitie Reading Tests, Level 5/6, Form K, Comprehension Section. During individual test administration, strategy use and error types were assessed using a protocol analysis technique through a combination of observation and retrospective reports. Actual strategy coding occurred at the time of the individual testing, while error coding was performed through analysis of audiotaped readings of all subjects subsequent to the individual testing.

Reliability of the experimenter’s error coding procedure and adherence to experimental procedures was obtained. The resultant proportional data was transformed using an arcsin transformation and analyzed using repeated measures ANOVAS. Appropriate post-hoc testing was performed when warranted, and precautions were taken to control for a positively biased F test.

The present study found support for differential responding by poor and average readers on the standardized, multiple-choice comprehension test. Reasons for students’ difficulties were found to be a combination of reading problems and test-taking strategy ineffectiveness. Although difficulties with word identification were found to negatively influence
comprehension test performance for both groups, it disproportionately affected the poor readers. The comprehension test administered seemed to more accurately measure the comprehension of average readers. The influence of meaning vocabulary errors was difficult to interpret in this study. However, it is certain that vocabulary difficulty also impacts a student's performance on a comprehension test, and that it is often confounded with word identification difficulties. That is, it is often difficult to assess the extent that vocabulary knowledge interferes with comprehension when students exhibit decoding failure, as indicated in this study.

The present study supports out-of-level testing (Fischer, 1962) for students who are suspected of having reading difficulty, if a more accurate reflection of a student's true comprehension ability is to be obtained.

Average readers and poor readers were found to use the same overall test-taking strategies, however with different response profiles. Furthermore, when average and poor readers were compared on their use of strategies when correct answers were obtained, the reasons for average readers' superior performances were clear. They used more positive strategies and used them more often. Although poor readers
used those same strategies, average students performed them more successfully. In addition, average readers were often observed to engage in immediate answer selection when a correct answer was obtained. Poor readers also used this strategy, however not as effectively.

When average readers and poor readers were compared on strategy use when incorrect answers were obtained, strategy 4 stood out as detrimental for both groups. That is, reading all the answer choices, but then looking back to match words, or guess, was ineffective, and often used when incorrect responses were obtained. Poor readers were found to use immediate answer selection more than average readers when incorrect answers were obtained.

These results supported the contention of certain previous researchers that poor readers as well as average readers definitely use specific test-taking strategies, but differ in their ability to regulate them. Average readers use better strategies more consistently, however, they are certainly not free from poor strategy use.

Results in this study supported interactive reading theories that recognize the effect of fluent, context-free word recognition as a fundamental component of reading comprehension. With the relative ease of word recognition on the part of the average readers, this allowed more
cognitive capacity for comprehension. Poor readers were not able to approach the comprehension stage to the same degree, since more attentional and capacity components were needed for word identification and were, therefore, not available to use for comprehension.

As an indication of self-monitoring behavior, students' propensity to alter answer selections was assessed. No differences between groups was found with the exception of poor readers changing answers from correct to incorrect more frequently than average readers. This indicates that both groups engaged in some self-monitoring, and were in fact trying to obtain correct answers. Possibly, poor readers are less certain of their answer choices and are more attracted to selection of obvious distractors.

Overall, it was shown in this research that students often exhibit a combination of reading difficulties and poor test-taking strategies when taking a multiple-choice comprehension test. Group comprehension tests are often used in schools as a measure of individual student comprehension level, however, group comprehension tests more accurately reflect the actual comprehension level of average readers than the level of poor readers. Despite this weakness, group comprehension tests seem to be
successful initially sorting out good readers from poor readers (although it may not be an accurate reflection of the comprehension level of poor readers).

In support of individual protocol analysis, observational data, and retrospective verbal reports, this method of individual testing was found to be informative and useful for gaining insight into the unspoken mental processes of a student during reading testing.
APPENDIX A

LETTER TO PARENTS AND CONSENT FORM
Dear Parents:

Our school has agreed to participate in a research project investigating reading skills of sixth-grade students. The research will be conducted by a certified school psychologist, Ms. Karen Anderson, from The Ohio State University, who has had experience testing and working with children. Your permission is needed for your child’s participation, please.

The purpose of the study is to discover the types of errors children make while taking group-administered reading comprehension tests. Information on why errors are made and what errors are commonly occurring, will help shed light on some of the problems children have while taking these types of tests. This information will benefit teachers in understanding your child’s reading strengths and weaknesses.

I need to work individually with your child for approximately one hour. (I need your permission for this, please!) Approximately 40 minutes will be devoted to reading several paragraphs. In addition, your child will also be administered a brief, nonverbal intelligence test (Matrix Analogies Test-Expanded Form) to obtain an estimate of his or her nonverbal ability. (Approximately 20 minutes).

The entire amount of time for individual work is one hour, and this will take place during school hours. I will be working in cooperation with two reading teachers, Mrs. Sunbury and Mrs. Millican.

Your child’s performance in the research will in no way influence school grades, since the information gained will only be used to help your child. This type of testing provides your child an opportunity to practice test-taking in a non-threatening situation.

If you have any questions regarding the study, please do not hesitate to call the experimenter, Karen Anderson, at 459-5344, or 457-0077. Please have your child return the attached page to Mrs. Millican or Mrs. Sunbury as soon as possible. Thank You!

Sincerely,

Karen Anderson, M. A.
Nationally Certified School Psychologist

Mr. _____________
School Principal
PARENT CONSENT FORM

I (circle one) do / do not give permission for my child (please print child’s full name), __________________________, to participate in the standardized reading test research program. I understand that all scores will be kept confidential, and that my child may withdraw from participation at any time without penalty.

Signature of Parent or Guardian __________________________ Date ______________

Name of Parent or Guardian (please print) __________________________ Relationship to child __________________________

Telephone number (DAY) __________________________ Telephone (EVENING) __________________________

IF YOU GIVE PERMISSION, please provide the following confidential information, which is necessary to select the children who will participate.

Is your child English speaking? Yes____ No____

Would your child be able to participate in the individual testing during the summer at your convenience? Yes____ No____

Since summer testing is necessary, when would be the best times for the examiner to contact you to schedule testing? (check all that apply)

_____ weekday mornings _____ weekday afternoons _____ weekday evenings

_____ weekend mornings _____ weekend afternoons _____ weekend evenings

PLEASE ANSWER THE QUESTIONS ON THE BACK OF THIS PAGE.
PLEASE RETURN THIS FORM TO YOUR CHILD’S TEACHER WITHIN THREE DAYS. THANK YOU!
Parents' or guardians' education (please check one in each column):

<table>
<thead>
<tr>
<th>Years Completed</th>
<th>Father or Male Guardian</th>
<th>Mother or Female Guardian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 8th Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th - 11th Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School Diploma or Equivalent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3 Years of College or Technical School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Years of College or More</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Check the one category that best describes each parent's occupation:

<table>
<thead>
<tr>
<th></th>
<th>Father or Male Guardian</th>
<th>Mother or Female Guardian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managerial, professional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical, sales, administrative support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farming, forestry, fishing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision production, craft, repair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator, fabricator, laborer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homemaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not currently in labor force</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Does this parent or guardian live with the child? (Please circle one in each column.)

<table>
<thead>
<tr>
<th></th>
<th>YES / NO</th>
<th>YES / NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father's or male guardian's occupation (please be specific):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother's or female guardian's occupation (please be specific):</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

May your child's school release recent test scores and grades for use in this research project?  ____ Yes  ____ No
# Individual Observation Sheet and Strategy Checklist

## Student ID

## Passage

### Item 1

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Observed or Reported</th>
<th>Observations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Read all answer choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Looked back in passage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Eliminated distractors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Matched words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Guessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Immediate selection</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations:**

**Answer Change:** yes no

### Item 2

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Observed or Reported</th>
<th>Observations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Read all answer choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Looked back in passage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Eliminated distractors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Matched words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Guessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations:**

**Answer Change:** yes no

### Item 3

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Observed or Reported</th>
<th>Observations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Read all answer choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Looked back in passage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Eliminated distractors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Matched words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Guessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations:**

**Answer Change:** yes no

### Item 4

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Observed or Reported</th>
<th>Observations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Read all answer choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Looked back in passage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Eliminated distractors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Matched words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Guessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations:**

**Answer Changes:** yes no
APPENDIX C

ERROR CODING SHEET
<table>
<thead>
<tr>
<th>Incorrect Item #</th>
<th>Self-Corr In Passage Tally</th>
<th>Word Indent/Word Reorg. Errors</th>
<th>Meaning Vocab Errors</th>
<th>Comprehension Errors</th>
<th>Comments/ Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewed Item</td>
<td>Adherence to Script</td>
<td>Read Passage Orally</td>
<td>Read Question Orally</td>
<td>Read Ans/choice Silently</td>
<td>Student explained ans/choice</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
</tr>
</tbody>
</table>
References


