THE RELATIONSHIPS AMONG ALTERNATIVE MEASURES OF READING COMPREHENSION IN LEARNING DISABLED STUDENTS

by

Lynn Z.Lang

Dissertation submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

in

Curriculum and Instruction

APPROVED:

Larry A. Harris, Co-Chairman	Barbara A. Hutson, Co-Chairman
Jerry Niles	Terry Wildman
Shirley Farrier	Mary Sigmond

December, 1983

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author wishes to express appreciation to the many people who made contributions to this study.

Special recognition is due to the co-chairmen of the writer's advisory committee, Dr. Larry A. Harris and Dr. Barbara Hutson. Dr. Harris' persistent encouragement during the development of the study made completion a reality, and without Dr. Hutson's conceptual guidance, the study would not have evolved as it has. Sincere thanks are expressed to the other members of the committee, Dr. Jerry Niles, Dr. Terry Wildman, and Dr. Shirley Farrier, for their participation and interest.

Thanks are also extended to whose sensitivity to the needs of handicapped learners in the public school setting has impacted greatly on the development of special education programs in the State of Virginia.

THANKS TO:

SPECIAL THANKS TO:

EXTRA SPECIAL THANKS TO:

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

CHAPTER

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

FIGURES Page

CHAPTER 1

Statement of the Problem

Purpose of the Study

The purpose of this study was to examine the nature of the relationships between and among the results of alternative measures of reading comprehension. In addition, the relative degree of support was evaluated for three possible models of relationship among alternative indicators of comprehension.

Introduction

Comprehending is the core of the reading process. Therefore measuring a reader's strengths and deficiencies in comprehending can become an indispensable teaching tool. In the broad sense, such a determination is made by "assessment," which is described in the literature as the "process of determining the behaviors . . . that a student does or does not possess" (Sedlak, Steppe-Jones, & Sedlack, 1982, p. 263) for the purpose of deciding how best to teach.

The assessment of reading comprehension behaviors in the classroom has not been an easy task. For the most part, this is so because the construct of comprehension is not clearly defined. Indeed, consensus among reading experts on a model of reading which defines and describes the interrelationships between the reader and all aspects of the

reading process, not just comprehending, is still not apparent (Farr, 1982). Until that consensus is reached, teacher knowledge of the best ways to measure and to develop comprehension in the classroom will be without a theoretical direction and focus.

There is evidence that gains are being made in understanding the nature of comprehension. New knowledge, resulting from a significant amount of research in reading over the last decade, is defining the comprehension process as a complex, multi-faceted process involving both linguistic/cognitive aspects of the reader and linguistic/logical factors of the text. In addition, a reader's own background of knowledge and experiences influences comprehension, but more important than any one aspect of the process is the interaction of these cognitive, linguistic, and prior knowledge factors in theoretical descriptions of how readers construct text meaning.

A foundation for reaching consensus about how comprehension should be defined, largely absent throughout the history of the reading field, is now being laid through research. The implications of such a consensus for the assessment of comprehension ability are significant.

Practical classroom application of knowledge about comprehension resulting from recent research has occurred to varying degrees, but the results of Durkin's (1978-1979) observations showed that the most frequently used strategy

for teaching comprehension in the classroom was asking students questions, also a frequently-used assessment procedure. Durkin's observations highlight the existence of an apparent gap between research findings and teachers' awareness of the complexities of the cognitive and linguistic process with which they are dealing on a daily basis. If full knowledge of these complexities were applied in classroom assessment and instructional routines, alternatives to questioning for the assessment as well as for the development of comprehension abilities in readers would be likely to be employed more frequently.

This study examines a variety of strategies for assessing reading comprehension that have been used in research and in the classroom during the last ten years. In view of the emerging description of comprehension, this study's first premise is that teachers may need to be more comprehensive than they have been in assessing readers' comprehension abilities. This means that assessment strategies may need to include a variety of close observations of those reader behaviors which contribute to comprehension's complexity, namely the "cognitive and linguistic activity that the reader displays" (Harste & Burke, 1978, p. 22). It also means that teachers may need to consider students' knowledge and experience backgrounds in interpreting comprehension assessment results. Essentially, it seems reasonable to assume that whatever assessment procedures are used in the

classroom, the assessment process should probably not be a simplistic one.

Second, this study operates on the premise that a teacher's conceptualization of comprehension is reflected in his/her choice of assessment strategies as well as of options for reading instruction (Jones, 1982). In practice, a teacher's narrow view of the complexities of comprehension typically results in limited opportunities for readers to express what they understand about text at any given time. If such a limitation occurs, teachers' existing conceptualizations need to be expanded so that the range of possibilities for allowing readers to reveal the operation and interaction of cognitive, linguistic, and prior knowledge factors in their comprehension will be increased.

Third, this study postulates that students should have a chance to demonstrate whether their comprehension is more or less than what is revealed in a standardized test score. This is a particularly important consideration since so many instructional decisions are based on standardized test performance.

Finally, this study is based on the position that research in the assessment of comprehension is necessarily tied directly to instruction. Specifically, as teacher knowledge of comprehension and appropriate assessment methodologies becomes more informed, particularly with respect to theoretical descriptions of the reading process,

instructional decisions are likely to be more responsive to individual differences in compehending abilities.

This chapter first summarizes some of the cognitive and linguistic aspects of comprehension currently identified in the literature upon which new conceptualizations and definitions of the process are now being based. Alternative strategies for assessing these various aspects of comprehension are then discussed.

Next, a rationale is given for measuring both cognitive and linguistic aspects of comprehension with two types of assessment: product measures of comprehension and process measures of comprehension. Essentially, use of these two types of measurement involves observing comprehension behaviors more closely and directly.

Next, possible patterns of relationship between and among the results of both product and process measures of comprehension are defined as three models of comprehension assessment. The issue of comprehension assessment with handicapped populations is also discussed briefly. Finally, the justification for and scope of the present study are summarized, followed by a listing of the research questions. The chapter closes with a statement of the study's limitations, definition of terms, and an overall summary.

Aspects of Comprehension

In 1970, Carroll wrote, "Comprehension may be regarded as a process that contains at least two stages: (a) apprehension of linguistic information, and (b) relating that information to a wider context" (p. 13). Since 1970, extensive research has confirmed both stages of the process as important components of comprehension. It has also been shown, however, that many additional factors affect a reader's ability to derive meaning from text.

Text characteristics, for example, have an impact on a reader's comprehension (Farr, 1982). Other factors, such as one's purpose for and interest in reading, the dynamics of the writer-reader interaction, the operation of reasoning and inferencing, and the reader's accumulated background and knowledge, also have a bearing on how much and what kind of comprehension results from reading.

A reader's background knowledge is affected by what he or she remembers (memory) as well as by the operation of schemata, theoretically described as cognitive organizers which structure the relationships between text information and the information represented by a reader's background and experience. For a review of the multiple factors affecting reading comprehension, see Singer and Ruddell (1976), Spiro, Bruce, and Brewer (1980), and Anderson and Spiro (1977).

In short, the complexity of comprehension appears to be one of the aspects of the process upon which most reading

experts now agree. This important generalization about the nature of comprehension necessarily complicates its assessment.

Product and Process Measures of Comprehension

One implication of the research on comprehension is that its multiple aspects might be indicative of multiple measurement techniques. In practice, teachers employ a number of options for measuring comprehension. In addition to asking questions, an informal comprehension assessment may include such well-documented measurement alternatives as free recall, free recall with probe questions, miscue analysis, and the cloze technique. It is not clear, however, that alternative techniques are regularly employed in measuring comprehension.

Teachers frequently assess students' competency in reading comprehension with standardized measures (Miller, 1978). Generally considered to be a formal measure of reading achievement, the standardized test is an efficient approach to comprehension assessment, but also one which does not allow teachers to evaluate directly how students manipulate language to generate meaning. Standardized testing is essentially a non-interactive assessment strategy from a teacher's viewpoint because it removes him or her from taking a role in making inferences about whether comprehension is occurring. The only role a teacher may have in the standardized test process is that of interpreting

judgments about comprehension based on guidelines presented in the administrator's manual.

When other comprehension assessments are used, one approach to interpreting their results is to view them as largely interchangeable. In this view, aside from small measurement differences or convenience factors, any comprehension measure should tap essentially the same knowledge as any other measure. While formal statements about the interchangeability of measures are rare, an implicit belief along these lines seems to underlie much of current school practice. If, in fact, various measures all indexed a behavior called comprehension, then it would make sense simply to choose the most convenient or most familiar measure for frequent use.

Another major approach to comprehension assessment is the evaluation of a student's separate mastery of each of a number of subskills of comprehension which have ostensibly been identified by reading experts as representative of the ability to understand text. This approach has been demonstrated in the diagnostic procedures associated with various subskill lists like those found in the Virginia Basic Learning Skills (1978) and the Virginia Standards of Learning (1981), as well as those found in some basal reader programs. The central assumption in this approach presumes that if a student achieves mastery of each subskill on a given list, then comprehension ability is ensured. Both the

standardized comprehension test and assessment of comprehension subskills are associated with the demand on educators to be more accountable for the results of instructional programs by demonstrating student achievement through test results.

Page and Vacca (1979) propose yet another approach to comprehension assessment which suggests that the diagnostic information resulting from the use of certain indicators of comprehension should be applied to corroborating the diagnostic information resulting from the use of other, but possibly different, indicators. To accomplish this, these authors suggest categorizing comprehension assessment alternatives on the basis of what they believe is a clearly researchable "distinction between responses elicited following reading (product indicators) and responses elicited during reading (in-process indicators)" (p. 54).

The basic differences between product and process measures of comprehension are that they measure a reader's use of language units of different sizes at different times in the reading process (Page & Vacca, 1979, p. 59). For example, answering questions about a passage and retelling passage meaning are product indicators of comprehension since both are administered <u>after</u> reading is completed, thereby measuring the product or result of reading.

On the other hand, in-process (or process) indicators of comprehension include cloze assessment and the analysis

of oral reading miscues. Both process measures require assessment at a different point in the process, namely the simultaneous observation of readers generating meaning while they are actually reading.

The logic for organizing comprehension assessment around Page and Vacca's product-process approach instead of around a standardized test, interchangeable measures, or a subskills approach is based on the need to evaluate comprehension during and after reading and to determine whether the results of alternative strategies corroborate each other or whether they show a variation in reader performance from product to process.

Johnston (1981) indicated that there are a number of reasons why a multiple-measures approach to comprehension assessment, similar to Page and Vacca's, is preferable to one which relies on a single test approach. In his opinion, the use of product as well as process indicators of comprehension represents a more flexible approach to assessment, provides more specific information about the reader, and finally, depending on the nature of a particular assessment task, provides a broader range of data about how a reader derives meaning from text (Johnston, 1981, p. 96).

Patterns of Association among Multiple Data Sources of Comprehension

Background. Research on measuring reading comprehension provides evidence that positive associations between

the results of alternative indicators of comprehension are often observable. Whether these observed associations mean that comprehension assessments are measuring a unitary factor, or are the result of a cumulative development of specific subskills of comprehension, or provide verification for the interaction of different components of the reading process that may be "not identical in function or strength but that are interdependent" (Guthrie, 1973, p. 10) is unresolved.

Based on a review of the comprehension assessment literature and on the way teachers typically assess reading comprehension in practice, three possible patterns of relationship between and among alternative indicators of comprehension (questioning, retelling, cloze, miscue) are proposed by this study. The specific nature of these patterns of relationship is defined as three possible models of comprehension assessment.

Assessment model one: high correlation among the results of four measures of reading comprehension (questioning, retelling, cloze, miscue). The first possible pattern of relationship is described as a high correlation among the results of all comprehension assessments. Model one is based on the assumption that all four assessments measure some common aspect of comprehension that is influential in determining the comprehension performance of readers, regardless of which assessment strategy is utilized.

If high relationships among the results of alternative measures are observed, then one would conclude that comprehension is influenced by a unitary factor and that its most efficient and effective measurement is use of a single test yielding an overall score which accurately represents the student's understanding.

For model one to be supported, the magnitude of all relationships observed among the results of the questioning, retelling, cloze, and miscue assessments for purposes of this study will be set at $r \geq .70$. This is, of course, an arbitrary value specified in order to provide a clear test of alternative models.

Assessment model two: low correlation among the results of four measures of reading comprehension (questions, retelling, cloze, miscue). A second possible pattern of relationship would involve a low correlation among the results of all comprehension assessments employed in this study. Model two is based on the assumption that there may be many different and distinct subprocesses or aspects of comprehension which can be identified (Davis, 1972) and measured.

If low relationships among the results of alternative measures are observed, then one would conclude that comprehension is a multi-dimensional phenomenon which requires multiple measurement of what may be discrete components or interdependent subprocesses.

For model two to be supported, the magnitude of all relationships observed among the results of the questioning, retelling, cloze, and miscue assessments for purposes of this study will be set at $r \leq .30$. Again, .30 is an arbitrary value specified in order to facilitate hypothesis testing.

Assessment model three: high correlation between the results of two product measures of comprehension (questioning, retelling)/high correlation between the results of two process measures of comprehension (cloze, miscue)/low correlation across product and process assessment types. A third possible pattern of relationship would involve higher correlations observed in the results of comprehension measures within assessment type (product; process) than are observed across both assessment types. Model three is based on the assumption that comprehension as product and comprehension as process require different types of assessment, and that each type of information contributes independently to assessment.

If correlations observed between the two measures within the product and within the process comprehension assessment category are higher than correlations observed across both categories, then one would conclude that the comprehension abilities required by product assessments (questioning, retelling) are similar. Likewise, a similarity in the reading behaviors required to demonstrate comprehension on process assessments (cloze, miscue) could also be assumed.

Finally, if correlations between the two types of assessment were low, one type could not be reliably substituted for the other.

For model three to be supported, the magnitude of an average of the correlations within assesment type for purposes of this study will need to be greater ($r \geq .70$) than an average of the correlations across assessments types ($r \leq .30$).

Comprehension Asessment of Identified Handicapped Students

The handicapping condition called "learning disability" or "specific learning disability" (SLD) has been the subject of much discussion over the past ten to fifteen years. Since SLD was officially recognized by the State of Virginia in 1970 as a handicapping condition for which special education services would be provided, programs to serve these students have increased dramatically.

In 1982, the Virginia State Department of Education reported that 38,604 identified learning disabled students between the ages of three and twenty-one were being served in Virginia as of December 1, 1982 (p. 5). This number was close to 40 percent of the total number of all students who were served in special education programs in the State of Virginia during that year. Enactment at the federal level of Public Law 94-142, the Education for All Handicapped Act of 1975, has insured free and appropriate public education for SLD students in the least restrictive environment.

Like the concept of comprehension, the nature of characteristics distinguishing a specific learning disability from other handicaps has been difficult to determine. Lerner (1976) identified factors such as neurological dysfunction or brain impairment, uneven growth patterns, a discrepancy between achievement and intelligence, and overall difficulty in academic and learning tasks as indicative of a specific learning disability. In practice, the gap between an individual's measured achievement and his or her intelligence remains a primary descriptor of the "classic" case of learning disability.

Despite problems in defining the handicapping condition, many students identified as SLD seem to have difficulty reading (Zigmond, Vallecorsa, & Leinhardt, 1982, p. 89). A selective review of the research in learning disabilities presented in Chapter Two demonstates how the field has shifted its major research emphasis away from investigating the perceptual aspects of reading (auditory and visual) toward examining the cognitive and linguistic demands of the process. In this regard, the learning disabilities literature reflects a greater compatibility with a cognitive view of reading than it has in past years.

A sample of SLD students participated in this study in order to conduct research in the area of reading and language with identified handicapped individuals. This was important for several reasons. First, research on the

comprehension abilities of SLD students was felt to be desirable because the decoding aspects of reading have been more frequently emphasized in the LD literature than has research on comprehension (Van Etten, 1978). Second, while the literature in the field reports research which examines the effects of alternative comprehension assessments, none of the studies address the specific differences between product and process comprehension assessments that are considered in this study. Third, the number of identified SLD students is increasing faster than any other category of handicapping condition. In view of this increase, a greater research effort must be directed toward the SLD population. Fourth, significant reading difficulties seem to be presented by learning disabled students. Hence, the examination of these students' performance under alternative reading comprehension assessment conditions represents an important opportunity to obtain information about how they process text information and demonstrate comprehension. Finally, the results of this study provide important empirical support for designing inservice activities for those special education teachers who are responsible for the instructional programs of SLD students.

Justification for the Study

The present study was needed to give further broad consideration to the question of comprehension assessment, and specific consideration to the question of how

interchangeable are alternative measures of reading comprehension. Since some have grouped alternative measures in the literature as product and process indicators of comprehension, this study was also needed to determine whether the distinction between these two types of measures would be apparent in the assessment data collected. A study of this nature has important implications for developing models of comprehension assessment as well as for refining comprehension assessment methodology in the classroom.

A number of researchers (Beebe, 1980; Carey, 1978; Carey, 1979, Fitzgerald & Fitzgerald, 1978; Goodman & Burke, 1973; Gould, 1979; Page, 1975; Sadoski, Page & Carey, 1981; Rigg & Taylor, 1979; Sadoski, 1981; and others) have taken a multiple-measures approach to comprehension assessment. Specifically, these researchers studied relationships among measures in an effort to learn more about the comprehension process. Overall, their findings support Page and Vacca's (1979) and Johnston's (1981) contention that a combination-of-measures assessment approach provides a better opportunity for making inferences about comprehension than does a single-test methodology.

The present investigation goes beyond other studies by examining the nature of relationships between multiple measures of comprehension that have been specifically identified as either product or as process assessments. In

addition, the present study has hypothesized and conducted an explicit test of possible patterns of association between and among these multiple measures of comprehension and equated the possible patterns of association to hypothesized models of comprehension assessment. Finally, this investigation examines the nature of relationships between multiple measures of comprehension in a sample of identified handicapped students.

Because the present study examines aspects of comprehension assessment that have not been fully incorporated into other studies, the researcher expects this study's findings to have implications for a refinement of theoretical models of comprehensions assessment as well as for the development of comprehension assessment methodolgy in practice.

Scope of the Study

The present study was designed to examine the nature of relationships between and among the results of two product measures of comprehension (questioning, retelling) and two process indicators of comprehension (cloze, miscue) in data collected from a sample of 48 male students who were identified as learning disabled.

In order to investigate the nature of observed relationships between and among product and process measures of comprehension, three possible patterns of association were hypothesized. Each pattern of association was proposed as a

model of reading comprehension assessment, distinguishable from other patterns by variation in the anticipated strength of association between questioning and retelling product assessments, between cloze and miscue process assessments, and among all four assessments in combination. Data collected from 48 seventh and eighth grade learning disabled males provided the basis for evaluating the relative support for the three proposed models of assessment in explaining the measurement of readers' comprehension.

While the study addressed the nature of relationships observed among product (questioning, retelling) and process (cloze, miscue) measures of comprehension, one external measure of comprehension was also used to permit comparisons with results of the four alternative assessments. The external measure of comprehension utilized was a standardized test of reading achievement. Standardized test scores were subsequently used as the dependent variable in a regression equation to determine which assessment procedures most closely agreed with subjects' comprehension performance as measured by the standardized test.

Research Questions

The following questions were addressed in the study's major analysis:

1. Does the observed data support any of the proposed models of comprehension assessment as described by three possible patterns of relationship in the results of

questioning, retelling, cloze, and miscue measures of comprehension including:

(a) High correlation among the results of four measures of reading comprehension (questioning, retelling, cloze, and miscue)?

(b) Low correlation among the results of four measures of reading comprehension (questioning, retelling, cloze, and miscue)?

(c) High correlation between the results of two product measures of comprehension (questioning, retelling)/ high correlation between the results of two process measures of comprehension (cloze, miscue)/low correlation among the results of product and process measures of reading comprehension?

2. Which reading comprehension assessments most closely agree with subjects' performance on a standardized group-administered reading comprehension test?

Limitations of the Study

1. The study used a correlational research procedure and, as such, was subject to certain restrictions in the analysis and interpretation of data. Because of these restrictions, the study's findings are only generalizable to groups similar to those from which the data were gathered.

2. Random selection of subjects was not possible due to severe limitations in the total number of students within the school district from which the sample was drawn who

might have met the specified study-entry criteria. This also limited the generalizability of the results.

3. One original criterion for entry into the study was that subjects' Full Scale IQ scores should place them within the average range of intelligence (90-109) as measured by the Wechsler Intelligence Scale for Children (Wechsler, 1974). Twenty-two subjects in the sample of 48 did not meet this criterion because the total population of students from which the sample was drawn was severely limited. This factor somewhat restricted control for subject variation due to intelligence.

Definition of Terms

1. Reading comprehension assessment - measurement of what a subject understands as indicated following reading by answering questions and retelling. In addition, reading comprehension assessment also refers to measurement of what a subject understands as indicated during reading by completing a cloze exercise and through the researcher's analysis of the semantic acceptability of oral reading miscues.

2. Product indicators of comprehension - questioning and retelling indicators of comprehension elicited following reading (Page & Vacca, 1979, p. 54).

3. Process indicators of comprehension - cloze and miscue indicators of comprehension, elicited during reading (Page & Vacca, 1979, p. 54).

4. Level of difficulty of test passages - the readability level of test passages used in the study, maintained between the 5.1-6.3 level of difficulty, as measured by the Flesch Readability Formula (1948) and the Gunning Readability Formula (1952).

5. Discourse analysis procedure - a method of dividing text into idea units of information (Taylor, 1978), based on conceptualizations articulated by Grimes (1975) and Meyer (1975).

6. Retelling scoring procedure - a method for determining whether or not the idea units of a passage are contained in a subject's retelling of a passage (Taylor, 1978, p. 11).

7. Learning disability - a handicapping condition as defined in Section 121a.5 of Public Law 94-142, the Education of All Handicapped Act of 1975.

8. Standardized reading achievement test - a standardized, formal, group-administered reading test which is "administered under specific conditions following directions given in an examiner's manual. Usually the time period allotted for completing such tests is also standardized" (Harris & Smith, 1972, p. 106). In this study, the Gates MacGinitie Reading Test, Level D, Form 3 (Houghton Mifflin, 1978) was utilized.

Summary of Chapter One

This chapter presented a justification for the present study in the context of the historical background of comprehension assessment, of contemporary descriptions of the comprehension process, and of current understandings about characteristics of learning disabled individuals. Research questions were stated and limitations of the study reviewed. This study was planned to contribute to the understanding of reading comprehension assessment methodology, particularly as it related to measuring the comprehension competencies of learning disabled subjects.

A review of the literature had disclosed no studies which examined the nature of relationships between and among four alternative indicators of comprehension, categorized as product and process assessments, on data collected from subjects identified as learning disabled. Neither had the literature reported research which tested specific models of comprehension assessment defined as possible patterns of relationship between alternative measures of reading comprehension.

The present study was motivated by the assumption that the assessment of reading comprehension may be unsatisfactory when teachers assume that a single comprehension test or simplistic assessment routine generates optimal data for planning instruction. A second assumption of the

investigation was that learning disabled students may actually be successful in comprehending text, but because of their specific learning problems, may not demonstrate their comprehension through traditional approaches to assessment, especially if only one test is used.

By observing the comprehension behaviors of handicapped students under a variety of assessment conditions, this study collected data useful in testing assumptions about what these students do when they try to gain meaning from text. In addition, the data represent an empirical basis for evaluating aspects of comprehension to the extent that they are revealed in product and process assessments. Finally, conclusions resulting from the analysis of data are potentially useful in designing teacher training activities which address comprehension assessment and its implication for instructional planning.

CHAPTER 2

Review of Literature

Introduction

This review of literature deals generally with the assessment of reading comprehension and specifically with the nature of relationships observed between and among alternative measures of comprehension.

Following a historical introduction, multiple data sources in comprehension assessment will first be discussed as they relate to different conceptualizations of the reading process. The remaining sections of the review will describe the task demands of two product measures of comprehension (questioning, retelling), two process measures of comprehension (cloze, miscue), and an external measure of comprehension, a standardized test. Since these measures are generally well-documented in the literature, a sample of their use in the research on assessing comprehension will be provided in this review. Subjects in the studies cited were drawn from both identified-handicapped and regular class populations.

Three models of reading comprehension assessment which hypothesize possible patterns of relationship both between and among product and process assessments will then be explained as they relate to different assumptions about comprehension appearing in the literature. Next, research relating to each of the hypothesized assessment models will be

reported. Finally, behavioral and cognitive characteristics of identified learning disabled students will be briefly reviewed in relation to emerging trends in the special education literature.

This review is illustrative in scope and focuses mainly on published journal literature dating from 1970 to the present. ERIC documents, Technical Reports from the Center for the Study of Reading, University of Illinois, and the resources of Dissertation Abstracts were primarily utilized in preparing the review.

Historical background. A long history of attempts by researchers, test designers, and teachers to find more valid and useful ways of assessing reading comprehension has been traced by Farr (1982) and Roser (1979). Farr reports that a "multitude of reading comprehension measures" (p.7) has been used over the years. Their validity and reliability are affected, however, by a lack of agreement among reading experts about what happens when a reader makes contact with text and tries to understand it.

As early as the 1930's, test writers incorporated the assessment of literal recall and vocabulary knowledge into published tests. In addition, they initiated the argument about whether reasoning abilities needed for inference drawing and interpreting were best assessed by reading tests or by measures of verbal intelligence.

Farr (1982) explained that most tests published in the 1940-1970 period were written as though reading comprehension were best described as a series of specific skills which could be identified, then precisely and discretely measured. A number of different subskill lists were indeed generated during that period, none of them developed according to a research-based "universally acceptable model of comprehension" (Farr, 1982, p. 8), however. Without this connection to a theoretical base, concerns about the validity of reading comprehension measures continue to be raised. Moreover, the appropriateness of instructional decisions based mainly on the results of invalid measures of comprehension is open to question (Roser, 1979).

Farr (1982) further reported that reading research has contributed to formulating the core of a theoretical model of reading within the last decade. As a result, it may be possible that a more comprehensive description of the various aspects of the reading process will lead to a more informed understanding of the phenomenon of comprehension itself, and thus to more valid assessment strategies.

Aspects of the reading process which research now suggests must be included in the assessment of comprehension include not only the assessment of a reader's use of language cues in getting meaning but also the assessment of such additional variables as reader background (memory, knowledge, experience), effects of text features, types of

inferences required to construct meaning, the reader's purpose for reading (interest, background), and the reader's application of information from text to practical situations (Farr, 1982). In addition, research applications of schema theory have resulted in even more information about how text knowledge and reader knowledge influence comprehension. Studies have examined how the reader's knowledge of syntax and grammatical complexity affect comprehension (Webber, 1980; Pearson, 1974-75), how the familiarity or unfamiliarity of text language alters comprehension (Tatham, 1970), how prior knowledge impacts on understanding text (Farr, 1982), how text structure and cohesion may affect recall (Marshall & Glock, 1978-79; Moe, 1979), and the newest area of research, how the reader's use of metacognitive strategies improves comprehension.

Despite the very considerable progress made in defining comprehension to date, questions continue to be asked and specific issues raised about the nature of the comprehension process itself, all of which have implications for the refinement of a valid comprehension assessment methodology. Some of these questions and issues will be addressed later in this chapter when the task demands of various comprehension assessment strategies are reviewed.

Multiple Data Sources of Reading Comprehension

Strategies for assessing comprehension differ from each other in a variety of ways. Some assessments are

administered only in a group setting while others require individual administration. Other variations in assessment include the oral or silent reading of text, requirement of a written or an oral response, an emphasis on processing small, medium, or large units of language during assessment, and assessment before, during, or immediately after reading (as well as on a delayed basis). In addition, comprehension assessments place different cognitive demands on readers, theoretically requiring them to become more or less involved in the construction of meaning. Finally, certain comprehension assessments tolerate more idiosyncratic responses from readers in their scoring systems than do other assessments.

Considering the historical absence of a unifying theoretical model of the reading process around which a valid comprehension assessment methodology could be designed (Farr, 1982), it is easy to understand why so many differences in measuring this complex behavior have persisted over time and continue to be observed in practice.

Assessment strategies used frequently during the 1970's reflect yet another difference: many are based on either a product or a process view of the reading process. First articulated by Simons (1971), the product-process distinction was a reflection of attempts by reading experts to more fully understand and describe the comprehension process.

In Simon's view, products of comprehension are observed by examining the information which a reader has gained as a

result of reading the text. This information is demonstrated by the reader in some outward fashion after the reading is over, usually by answering questions or by retelling story content. Assessing the products of reading is supported by a text-based view of comprehension which holds that exact reproduction of the author's meaning should be one of the major indicators of effective comprehension. It is an approach to assessment which leaves little room for much reader in-put in the effort to comprehend.

Processes of comprehension, on the other hand, are inferred from the active involvement of the reader in gaining information from text, in using text information to construct meaning, and in relating text information to one's knowledge and experience base. Assessing the processes readers use to construct meaning is supported by a language-based view of comprehension and involves collecting data from readers while they are "in the process" of reading. Observing comprehension in process may include such vehicles as the analysis of oral reading miscues and the cloze procedure.

In addition to text-based (product) and language-based (process) views of reading, another view of reading assumes that both the products and the processes of comprehension are important to understanding text. Theoretically, this view holds that neither aspect can be excluded from describing the comprehension process and that each one must be

incorporated with information coming from the reader's experiences and knowledge background to represent the construct fully. Rumelhart (1976) represented this simultaneous integration of text-based, language based and knowledge-based information as an interactive model of the reading process. Assessment based on an interactive view of reading would necessarily involve the use of both product and process indicators of comprehension.

In summary, elements of the text (such as print features, text structure, cohesion) account for one aspect of reading which affects readers' comprehension. In addition, readers' accumulated knowledge of the subsystems of language (including grapho-phonic relationships, and schema for syntax and semantics) affects comprehension, along with one's background of knowledge and experience.

In the section that follows, the assessment task demands of four alternative measures of reading comprehension (questioning, retelling, cloze, and miscue) will be discussed. In addition, characteristics of standardized test procedures will be briefly reviewed. For each assessment strategy discussed, a definition of its task demands will be offered, followed by an illustrative application of its use in research, and concluding with an indication of problems typically associated with the measure.

Task Demands of Product Assessments

Introduction. Page and Vacca (1979) discussed at length the benefits of assessing comprehension with both product and process measures. They felt it represented an approach to assessment which was not only more comprehensive in scope, but which also satisfied the requirements of reading experts who valued the use of both qualitative and quantitative measures of comprehension (Hood & Kendall, 1979), who believed that students' behavior should be observed while they are engaged in reading (Farr, 1970), and who wanted comprehension to be assessed in both silent and oral contexts (Newman, 1978).

Product indicators of comprehension, according to Page and Vacca (1979), "emphasize how much the reader has learned as a result of comprehending passage content" (p. 51). In this study, answering questions and producing an oral retelling are the two product indicators of comprehension under investigation. These measures are called product indicators because they examine the results of reading, or that exact reproduction of the text which a reader can demonstrate after a certain amount of reading has been done.

Questioning. There is little disagreement that questioning, either orally or written, is a predominant comprehension assessment technique. Pearson and Johnson (1978) write, "Questions have been the mainstay of reading comprehension instruction for decades. They appear on most

standardized and informal comprehension tests. Workbooks are filled with them, and teachers' manuals provide numerous prereading and postreading questions to use in discussing the selections that students read" (p. 154).

One concern about questions frequently addressed by the literature is what kind of questions should be asked. This issue merits consideration because of the assumption that a reader's responses to questions may vary according to the type of question that he or she is required to answer. Questions are generally classified as either literal, interpretive, or evaluative in nature, theoretically constructed that way in order to assess and sometimes to stimulate readers' thinking about text at increasingly higher cognitive levels. Different question types may also be equated to different levels of reader comprehension skills as specified on basic skills lists. For example, the comprehension behavior of recognizing and recalling facts parallels literal level questions, inferring beyond the facts parallels interpretive level questions, and evaluating what is read against certain specified criteria parallels evaluative level questions (Guszak, 1969; N.B. Smith, 1969).

Guszak (1967) reported that teachers' most frequently used questions are formed at the literal level, directed toward obtaining "who, what, when, where, how" responses from readers with little emphasis on the interpretive level of comprehension. Harste & Burke (1978) concluded that such

literal questions force readers to focus too heavily on text information and tend to assess how well they can represent exactly what the author wrote. An answer to literal or lower-level questions is usually correct or incorrect. The same standard of correctness cannot always be maintained with questions that seek interpretive or evaluative responses, however.

In research relating to question-type, C.T. Smith (1978) studied both the length and grammatical complexity of the oral responses of 30 second graders and 30 fourth graders to both factual and interpretive questions. She found that all subjects responded in significantly longer communication units when interpretive level questions were asked (p. 898). Smith considered this finding particularly important in light of language research which suggests that the level of children's cognitive development can be inferred by examining the length and complexity of their oral language productions. Smith's study concluded that higher level cognitive functioning may, in fact, be stimulated by asking higher level questions.

Lovitt and Hansen (1976) conducted research on question type as a factor in the reading comprehension of a small sample (N=7) of identified learning disabled boys between nine and twelve years of age. Collecting data over one academic year, the authors found that students' answers to written comprehension questions following daily oral and

silent reading were more often correct at the recall level. In fact, recall questions were answered correctly more frequently than were questions involving the use of sequencing and interpreting abilities. While a consistently applied intervention technique was effective in increasing the number of questions answered correctly at all three levels, subjects continued to produce more correct responses to recall level questions over the course of the year. The results of this study confirm the widely held assumption that recall level questions are the easiest to answer of any question category, and yet a sample of identified handicapped subjects showed improvement in their ability to answer even interpretive level and sequencing questions with training.

Both the Smith and the Lovitt and Hansen study tend to support Durkin's (1980) assertion, in a review of the research implications of questioning, that a questioner's skill in asking questions of different types may be an important factor in the effectiveness of the methodology as a diagnostic and instructional strategy.

Notwithstanding question type, Newman (1978) concluded that "higher level skills . . . are best evaluated when the reader has opportunities for careful, thoughtful contemplation" (p. 885). He wrote that responding to questions following silent reading provides a better opportunity for readers to "interpret, visualize, analyze, synthesize,

evaluate and apply what s/he has learned from print . . ." (p. 885) than does the recounting of passage meaning following oral reading.

A second issue under discussion in the literature is that of question placement. Durkin's (1980) discussion of the literature relating to questioning reviewed research which showed that asking questions before and after reading increases students' learning from text by focusing their attention more directly on (a) text content that relates to a specific question, in the case of pre-posed questions, or (b) text content that relates to specific questions as well as to broader understandings in the text, in the case of post-posed questions (p. 30). Singer (1978) shared Durkin's opinion that both pre-posed and post-posed questions help students organize their thinking and focus their attention on particular information within text that relates to the question, thus facilitating overall passage comprehension.

Regardless of the kind of questions asked or when they are asked in relation to reading, the literature identifies numerous shortcomings associated with this assessment strategy and addresses them in detail.

The first problem raised is the extent to which one's question-answering ability may be confounded by memory. After conducting extensive observations in each of 36 third-to-sixth grade classrooms of so-called "best teachers," Durkin (1978-79) reported that teachers ask many questions

during reading and social studies lessons. In addition, questions may be asked quite a long time after the actual text reading occurs, sometimes days later. If these findings are representative of actual classroom practice, then it seems likely that questions may be assessing one's memory of what was comprehended at some earlier time, rather than the actual comprehension of text information itself.

A second problem with the assessment strategy is that a reader's answers to questions may be cued by the very words of the question itself (Hansen, 1978; Marr & Gormley, 1982; Page & Vacca, 1979; Singer, 1978). Marr and Gormley (1982) suggest that a desire to eliminate answer cues in the language of questions may have been one reason for selecting the retelling format to assess comprehension in the Reading Miscue Inventory (Goodman & Burke, 1972). The Goodman and Burke (1972) retelling format involves use of probe questions, which are asked only in relation to actual information appearing in the retelling (Marr and Gormley, 1982, p. 91), thereby minimizing their answer-cueing potential.

Recently, taxonomies of questioning reflecting some of the more recent conceptualizations of comprehension have been described. They include questions which acknowledge the reader's contribution to the comprehension process not only through one's knowledge of language but also through a knowledge and experience background. In Pearson and Johnson's (1978) taxonomy of questions (discussed more fully

in Chapter Three), a reader's comprehension can be tapped at both explicit and implicit levels, requiring that explicit and implicit text information be used to generate answers to questions. In addition, comprehension can be tapped at the scriptal level, requiring that answers to questions be based solely on information drawn from the reader's accumulated knowledge and experience base.

In summary, asking questions about passage content is a frequently used approach to assessing reading comprehension. While research has shown that questions can be structured to engage a reader in higher levels of cognitive activity, many of them do not. Nevertheless, asking questions of different types and at different times in relation to reading has effectively increased students' comprehension of text information by focusing their attention and by providing structure for organizing their thinking about what is read. Finally, questioning techniques are now reflecting new information about the importance of allowing the reader's prior knowledge and his/her purposes for reading to enter into the assessment process.

Retelling. The retelling assessment strategy demands much of the reader and has been subjected to critical review by reading experts. Most of the criticism centers around the fact that retelling protocols are difficult to score with complete objectivity. Likewise, because the organizational demands of the retelling task are great, a reader may

tend not to retell all of the information that he or she has comprehended.

In order to retell a reader must not only organize text information but must also relate that information to a listener through the vehicle of oral language. Johnston (1983) listed additional choices a reader must make to produce a retelling including how much detail should be included in the retelling, where the retelling should start, and in what order subsequent events should be retold.

All of the behaviors required for retelling make it an assessment task which requires a level of reader cognitive activity and involvement in assessment that is not found in most other indicators of comprehension. Rigg and Taylor (1979) address this difference between retelling and other assessments in their conclusion that, "The CAT (California Achievement Test) and cloze measures what readers can produce that satisfies the requirements of the questioner; retelling measures what the readers produce that satisfies themselves. This is a difference that cannot be emphasized too strongly; unlike other tests, retelling asks the reader to establish the criteria of importance and completeness" (p. 82).

Assessment of a reader's recall of text information was one of the earliest measures of reading ability used. In his history of comprehension assessment, Farr (1982) reported that as early as 1913, Rudolf Pintner measured the

comprehension of fourth-grade subjects by asking them to write down everything they could remember about a passage. Likewise, Durrell used an oral procedure to assess comprehension in the silent reading subtest of his Durrell Analysis of Reading Difficulty (1955) whereby students "read and then recited orally from memory" (Farr, 1982, p.4).

In an essay on retelling as a measure of comprehension performance, S. Smith (1979) wrote that whether retelling is done in an oral or in a written format, it is first and foremost an "open response verbalization" (p. 90), meaning that no advance information is supplied to the reader about what may be important to remember. In this sense, Smith considers the retelling task to be "the (purest) form of reader response" (p. 90), and certainly different from answering questions.

Research using retelling as an indicator of reading comprehension reports findings which tend to confirm the uniqueness of the strategy as an assessment device. Some of the key outcomes of this research are summarized below. First, there are indications that readers actually retell very little information in comparision to the quantity that is available for retelling. In a study subsequently discussed in greater detail, Tierney, Bridge and Cera (1978-79) reported that even good readers sometimes retell as few as one-half of the total items identified for retelling. Meyer's (1977) study, also discussed later, reported that

readers recalled only a small portion of the total number of possible ideas in test passages used for data collection. In this instance, they recalled only 23 percent of the total possible information.

The findings of both these studies tend to confirm the assumption that a reader's attempt to first decide what should be retold, then organize it in some "manageable form" for presentation might be adversely affected by limitations in memory, as well as by other constraining factors inherent in the demands of the task itself (Tierney, Bridge, and Cera, 1978-79, p. 552).

A second conclusion about retelling drawn from the research is that many readers include more main idea text information in their retellings than details. Meyer (1977) studied the retelling patterns of 23 college students who had not been grouped by ability and compared the information they retold to a content structure grammar. She found that subjects retold more items listed near the top of the content structure outline in both immediate and delayed retelling situations, regardless of whether they had been exposed to the story one, two, or three times. Further, recalled items tended to be in close proximity to each other on the hierarchical outlines. As a result of these findings, Meyer hypothesized that readers remember main ideas more easily and, in fact, may use them as a structure for remembering text details.

In another study, Dunn, Matthews, and Bieger (1979) used a written retelling format to compare the retellings of a group of 92 fourth and sixth grade readers to a hierarchically arranged content structure outline. Comparative analysis showed that the better readers in the group included proportionately more main ideas (superordinate) as well as detail ideas (subordinate) in their retellings than less skilled readers did, confirming Tierney, Bridge, and Cera's (1978-79) findings, reported later in this section. In addition, the less skilled readers exhibited a pattern of "decreasing recall with increased subordination of information in the content structure" (p. 20). Dunn, Matthews, and Bieger (1979) hypothesized the existence of developmental differences in the relative cognitive abilities of good and poor readers, as demonstrated particularly by the poor reader's limited capacity to understand the more complex semantic relationships of language that are frequently expressed in the subordinate ideas of text (pp. 20-22).

Documentation of the differences observed in good and poor readers' retellings represents a third area of inquiry which has been addressed by the research. Overall the content of retelling protocols seems to vary in relation to the reading ability of subjects. Tierney, Bridge and Cera (1978-79) for example, compared the retellings of 36 third graders, 18 of whom were classified as good readers and 18 classified as poor readers, and found that good readers

included more information in their retellings, overall. This was true for both explicit and inferred information related in unaided retellings or related in response to probe questions. In addition, both good and poor readers retold more inferred items when probe questions were asked than in the unaided situation. This tendency to retell inferred items when probe questions were asked seemed to be particularly strong among good readers and shows how one assessment strategy can strengthen the results of another.

In another study of good and poor readers, Taylor (1975) found that the comprehension of 31 fifth grade poor readers as assessed by retelling was greater than that demonstrated by the same number of third grade good readers when both groups read familiar text content at the third grade level of difficulty. The poor readers' total retelling scores on the easiest, familiar content passage were superior to the younger good readers' scores because the poor readers retold slightly more passage details than main ideas. These findings suggest that poor readers can meet the cognitive and production demands of the retelling task, especially when text content is familiar to them and when decoding requirements are manageable. The fifth grade poor readers' inclusion of more details in their retellings of the easy passage might have been accounted for by their greater background and experience in comparison to the younger students. Their strength was not maintained when

text content was unfamiliar and written at a higher reading level, however. While Taylor's results reinforce the usefulness of retelling as an assessment of comprehension even with poor readers, they also tend to confirm the existence of basic differences in the capacities of good and poor readers to understand text.

A fourth area of basic interest in the literature is whether information presented in retellings can be maintained over time and then retold later. Rigg and Taylor (1979) studied a sample of fifth graders to find out how the retellings of 12 subjects who retold immediately after reading (one retelling-immediate only) and immediately after reading plus one week later (two retellings-immediate and delayed) differed from the retellings of 12 subjects who retold only on a delayed basis, one week after reading the text. Both delayed retelling groups were informed in advance that they would be asked to retell in a week's time. Analysis showed that the retelling scores of the "immediate plus delayed retelling" group were significantly higher than were scores for the immediate-only and delayed-only groups. Whether the rehearsal aspect of an immediate retelling followed by a second retelling in one week's time actually caused the score differences between the two groups was not addressed, but the repeated retellings did appear to have a positive effect on comprehension. Along similar lines, when Meyer's (1977) research showed that subjects remembered and retold

more main ideas, she concluded that they used these ideas as a structure for remembering details, and that this relationship could be observed in both immediate and delayed retelling situations (p. 183). While there is some evidence that the ability to retell text information is maintained over time and that practicing retelling may be beneficial, the instructional potential of these findings has not yet been fully explored.

Since receiving attention as part of the diagnostic procedures of the Reading Miscue Inventory (Y. Goodman and Burke, 1972), the difficulties in scoring and interpreting retelling protocols as well as the potential usefulness of retelling information as diagnostic data have been discussed frequently in the literature.

Criticism has been primarily directed toward difficulties in achieving consistency among diagnosticians in scoring and interpreting retell protocols. Johnston (1983) suggests that recent advances made in the use of text content structure outlines as a basis for achieving greater reliability in scoring retellings have application both in research and in classroom settings. Indeed, content outlines were used as guidelines for scoring retellings in many of the studies reported in this section.

With regard to the usefulness of the retelling strategy as an indicator of a reader's "retention of meaning" (Goodman & Goodman, 1977, p. 320), most concern centers

around the issue of how to treat text information that is not retold. If a reader produces a superficial retelling, for example, should the assumption be made that it represents all that was comprehended, or that he or she might have been able to retell more? Johnston (1983) believes that information may be omitted from retellings for a number of reasons. It can happen because a reader misunderstands the requirements of the retelling task, because he or she thinks the teacher already knows the details of the passage, and/or because the actual production demands of the retelling task limit completeness. Royer, Hastings, and Cook (1979) believe that variability in subjects' retelling protocols may even be associated with motivational and emotional factors, such as characteristic shyness in a reader's personality.

On the contrary, S. Smith (1979) accepts the fact that readers are not likely to retell all the information they gain as a result of reading. She feels that a teacher should be more interested in "how the student is thinking about the content" (p. 90) rather than how much of it is actually retold. Goodman and Goodman (1977) concur with this view and suggest that readers' organization of text information, the presence or absence of the author's or their own language in retellings, and evidence of conceptions and misconceptions about text content in retellings all provide diagnostic insights which are useful in making

inferences about the cognitive activity that may be a part of comprehending (p. 320). Likewise, Harste and Burke (1978) write that observing the "cognitive and linguistic activity that the reader displays" during reading and retelling (p. 22) represents the primary diagnostic potential of the assessment strategy. For an interesting discussion of how inferences are revealed in retellings, see Gould (1979).

In summary, reading experts tend to agree that the cognitive demands of retelling on the reader are significant, but that the scoring and interpretation demands on the teacher complicate the evaluation of potentially powerful information about the cognitive aspects of a reader's search for meaning. Research findings indicate that subjects usually retell only a small percentage of the total items identified for retelling. Whether this is best explained by the memory and production requirements of the task or by other factors is uncertain. Despite limitations in completeness, differences between the retellings of good and poor readers are apparent, with good readers generally supplying more of all types of information (explicit, inferred) in their retell protocols than poor readers. Furthermore, while there is evidence that main ideas may be more easily recalled, poor readers seem to retell proportionately fewer idea units overall, but are able to retell when the difficulty level of text is not overwhelming.

Standardized reading achievement tests. A standardized measure of reading comprehension was used in this study as an external indicator of comprehension performance. It is discussed at this point in the review of literature because standardized test results are generally considered outcomes or products of reading. In most standardized comprehension tests, a student reads a number of short selections and answers questions about each one when reading is finished. For a comprehensive description of the true-false, multiple choice, and fill-in-the-blank questions found on most survey type, formal measures of reading achievement, see Johnston (1983, pp. 58-61).

The results of standardized tests are most frequently used as a data base to support educators' "selection and classification" decisions (Johnston, 1981). Pearson and Johnson (1978) call this kind of testing "comparative assessment" and suggest that it is most applicable to educational situations where a record of student progress by pre-post tests is needed, where a basis for assigning students to a particular group or level is needed, and where data representing the accountability of an instructional program is required by school district officials or by the public.

Standardized tests administered to select and to classify students have a certain utility. One problem with their use in this context, however, is that the same basis

of comparison is employed for all students in a group. As a result, tests must necessarily contain a wide-enough range of items to discriminate adequately between the variety of ability levels represented in the group, thus increasing test length to a point of unmanageability.

Standardized test results are useful in a second situation. They can provide necessary support data for making certain "administrative decisions" regarding the success or non-success of a particular educational program, piece of instructional material, or teaching method (Pearson and Johnson, 1978). Since many of these decisions involve the continued expenditure or withdrawal of school district, state, or Federal financial resources, standardized test data are important and valued components of the decision-making process.

One area where use of standardized test data is inadequate, however, is in making instructional decisions regarding the specific educational program of individual students. Percentile scores and the more confusing and deceptive grade equivalent scores do provide teachers with a global indication of a student's reading behavior, but they also offer little specific information about what a student actually does when he or she reads, and thus, cannot be relied on for specific instructional planning. Even though Farr (1982) recommends that teachers use standardized tests as "one piece of information that recommends on-going

assessment" (p. 32), many experts still support an approach to assessment which directs greater teacher attention to "how individual children process printed material" (Page and Pinnell, 1979, p. 81).

There are a number of particular problems associated with using standardized tests to assess comprehension, the first being that teachers may use them to the exclusion of most other measures. This is likely to happen because standardized tests are relatively easy to administer and score.

A second problem inherent in standardized tests is their length, especially with regard to the number of test items needed to measure a complex behavior like comprehension. Given the continued refinement of knowledge about what constitutes comprehension, it is possible that test length might need to be substantially increased in order to assess reliably all of its various aspects (Pearson and Johnson, 1978). Johnston (1983) writes that total reliance on standardized test data is "futile" (p. 54), particularly in view of the "complexity of the reading task and the number of variables to be assessed and/or taken into account" (p. 54).

A third problem with standardized tests of reading comprehension involves reaching agreement about what behaviors they actually do measure. Carroll (1977) suggested that tasks found on reading tests are not only related to the measurement of reading comprehension abilities but also

to the measurement of language comprehension and cognitive abilities. He further stated that existing tests have provided no way to compare a students' reading comprehension, language comprehension, and cognitive functioning. As a solution, Carroll proposed a strategy which would measure a student's comprehension of print material and then compare it to an assessment of his/her oral comprehension, without a print component. In addition, a separate evaluation of cognitive or conceptual ability would be done to "index the relative levels of a reader's ability to (a) handle the conceptual and inferential reasoning processes involved in reading, (b) handle the language comprehension elements in reading, and (c) do these things in the print medium as well as in other media" (Carroll, 1977, p. 9).

A fourth problem generally attributed to standardized comprehension measures is the extent to which test items can be answered without reading the accompanying test passages (Tuinman, 1973-74). When a reader must refer to test passages in order to make a response, the test items are considered to be passage dependent. This is a positive characteristic and an indication of a test's reliability as a standardized measure.

Finally, the existence of cultural bias in standardized measures may limit the performance of students whose cultural backgrounds and experiences do not match those expressed by test items. Such differences need to be considered when

evaluating students' performance. Pearson and Johnson (1978) conclude that, "test publishers are attempting to remove as much cultural bias (or trying to balance it) from their tests as they can while avoiding the undesirable situation of not requiring students to relate textual information to prior experience in order to answer test items" (pp. 212-213).

In summary, standardized measures of reading comprehension are frequently used in the educational setting. While they have limited usefulness and application in providing data support for selection, classification and administrative decisions, their use presents a number of problems. In addition, their results offer a relatively narrow picture of the complexities of comprehension revealed in a reader's behavior. In order to gather more specific information about students for instructional planning purposes, measures of reading performance which are more informal in structure, administration and scoring are required. Such measures have the potential for allowing teachers to observe the reader more closely and to examine some of the "processes with which readers operate" (Kemp, 1976, p. 82).

Task Demands of Process Assessments

Introduction. Process indicators of comprehension, according to Page and Vacca (1979), are represented by responses that demonstrate a reader's comprehension "elicited during reading" (p. 50). In this study, data resulting from

the completion of a cloze exercise and from the miscue analysis of oral reading errors are the two process indicators of comprehension under investigation. These measures are called process indicators of comprehension because they generate data which serve as the basis for inferring what cognitive and linguistic processes a reader may be using to comprehend.

Cloze. Cloze methodology was first articulated by Taylor (1954) who derived it from the gestalt concept of "closure." Now used frequently as a comprehension assessment strategy, a measure of readability, and an instructional technique, cloze requires a reader to produce meaningful "closures" by filling in words that have been systematically deleted from text.

Cloze is based on a belief that "the task of filling deleted words in a written text is related to the way readers process language in reading" by using "knowledge of language and his/her prior experience with the content of the material to anticipate and reconstruct the meaning of the message" (Kaminsky, 1979, p. 13). Generally considered an indicator of "intra- rather than intersentence comprehension" (White, Pascarella, & Pflaum, 1981, p. 703), cloze methodology assumes that "how the reader controls the author's lexicon relates directly to control of deep structures and semantic meaning" (Harste & Burke, 1978, p. 16). For a discussion of cloze construction, of variations in

deletion patterns, and of different cloze scoring criteria, see Heerman & Treadway (1978), Kaminsky (1979), Rankin and Culhane (1969), and Karlin (1973).

Despite the rationale which verifies cloze as an indicator of reading comprehension, abundant literature on the methodology suggests that its emphasis on manipulating the surface structure of language as well as its frequent interruption of the reading process by blank spaces in text "may well measure some process quite unlike the reading process as it might occur under natural conditions" (Harste & Burke, 1978, pp. 16-17). Page and Vacca (1979) and Kaminsky (1979) concur with this view. Kaminsky writes, "In fact, the focus on individual words which the cloze task requires is frequently in opposition to the way in which the successful reader might eliminate irrelevant words as he/she is reading" (1979, p. 17).

Further, the cloze task's word emphasis has been found to be particularly difficult for young readers who are successful in filling in gaps at the oral level, but whose limited skills in processing print language (decoding skills, tendency to read word-by-word) are not strong enough to cope with the additional confusion caused by a mutilated text. Deck (1979) suggested that if cloze were to be used effectively with a young group, the "semantic constraints implied by context must be rather explicit" (1977, p. 14).

Other problems with cloze methodology include its anxiety-producing nature (Miller, 1978), its insensitivity to distinguishing readers who generate high quality synonyms for cloze answers from readers who give totally inappropriate answers (Kaminsky, 1979), and "the confusion which occurs when deleted words are linguistically random" (Kaminsky, 1979, p. 16).

Even though cloze methodology has been the subject of much critical review by some experts in the field, it has also been shown to have a number of advantages in testing and teaching comprehension. Regarding assessment, a cloze test is easily constructed using classroom reading materials, is easily administered to a large group, and is easily scored. Likewise, "examiner intrusion" (Russell, 1979) is minimized because the tester has no role in deciding which words should be deleted, nor does the assessment rely on an examiner's questions, thus avoiding the problem of determining whether readers may have misunderstood the questions rather than the passage content.

High correlations between cloze and standardized indicators of comprehension (see Table 1) have caused cloze to be regarded as a measure of the same skills that are measured by standardized multiple-choice reading texts (Bormuth, 1969; Rankin & Culhane, 1969) and in that sense, cloze is thought to be a measure of "essentially . . . literal level" (Kaminsky, 1977) reading comprehension skills. Results of

Deno, Mirkin & Chiang's (1982) studies, discussed at a later point in this chapter, found that cloze performance appeared to be more highly associated with word recognition skills than with comprehension skills as measured by the standardized test (p. 93). Regardless of which aspect of the reading process cloze has been most directly associated with, it has been shown to be effectively used with students whose native language was not English (Baldauf, Dawson, Prior & Propst, 1980), and with retarded students (James, 1975; Hargis, 1972), even though Hargis found that retarded students' restricted range of vocabulary made the production of synonyms for cloze blanks a difficult task (p. 728).

As a testing strategy, cloze has also been established as an indicator of text readability by the relationship of specific cloze scores to independent (50 percent or above correct), instructional (40 to 50 percent correct), and frustration (below 40 percent correct) reading levels (Bormuth, 1968). Ransom's (1968) correlation of elementary students' scores on a cloze measure and on an informal reading inventory further confirmed the usefulness of cloze test results in reflecting passage readability.

As a teaching strategy, cloze has been found to be a valid technique for helping readers recognize the "relational elements in written material . . . and syntactic units" (Seidenberg, 1982, p. 353). Seidenberg suggested that since learning disabled students may not have developed the

TABLE 1
Relationships among the Results of Multiple Data Sources of Comprehension

Author	Date	Assessments Utilized	Relationship(s) Reported	N
Deno, Mirkin and Chiang	1982	STUDY II		
		Isolated Words and underlined words in Context	r=.88*	45
		Isolated Words and Oral Reading of Passage	r=.92	
		Isolated Words and Cloze	r=.84	
		Isolated Words and Word Meaning	r=.60	
		Underlined Words in Context and Oral Reading of Passage	r=.88	
		Underlined Words in Context and Cloze	r=.81	
		Underlined Words in Context and Word Meaning	r=.72	
		Oral Reading of Passage and Cloze	r=.86	
		Oral Reading of Passage and Word Meaning	r=.57	
		Cloze and Word Meaning	r=.50	
		STUDY III		
		Oral Reading of Passage and Literal Comp. Subtest of Standardized Test	r=.78	
		Oral Reading of Passage and Inferential Comprehension Subtest of Standardized Test	r=.80	
		Cloze and Literal Comprehension Subtest	r=.67	
		Cloze and Inferential Comprehension Subtest	r=.71	
J. Guthrie	1973	Phonics Test Production Subtest Intercorrelations for Disabled Readers:		48
		Nonsense word product and long vowel production	r=.50	
		Nonsense word product and short vowel production	r=.85	
		Nonsense word product and consonant cluster production	r=.69	

* Based on third grade level materials and a 60-second time limit for reading.

TABLE 1 (continued)
Relationships among the Results of Multiple Data Sources of Comprehension

Author	Date	Assessments Utilized	Relationship(s) Reported	N
J. Guthrie	1973	Nonsense word production and Single Letter Production	r=.37	
		Long Vowel Production and Short Vowel Production	r=.52	
		Long Vowel Production and Consonant Vowel Production	r=.34	
		Long Vowel Production and Single Letter Production	r=.18	
		Short Vowel Production and Consonant Cluster Production	r=.66	
		Short Vowel Production and Single Letter Production	r=.40	
		Consonant Cluster Production and Single Letter Production	r=.49	
C.L. Hansen	1978	1) Grade 3 materials		
		Propositions Recalled and Questions Answered (LD students)	r=.61	17
		Propositions Recalled and Questions Answered (Regular students)	r=.77	17
		2) Grade 5 materials		
		Propositions Recalled and Questions Answered (LD students)	r=.46	17
		Propositions Recalled and Questions Answered (Regular students)	r=.50	17
Adams, Carine, and Gersten	1982	1) Retelling and Questions Answered (immediate)	r=.32	45
		2) Retelling and Questions Answered (delayed)	r=.57	45

TABLE 1 (continued)
Relationships among the Results of Multiple Data Sources of Comprehension

Author	Date	Assessments Utilized	Relationship(s) Reported	N
Carey	1978	Miscue Comprehending Score and Post-Oral Cloze Test	r=.55	100
Page	1977	Miscue Comprehending Score and Post-Oral Cloze	r=.43	48
Sadoski	1980	Miscue Comprehending Process Score and Post-Oral Cloze	r=.55	48
Goodman and Burke	1973	Miscue Comprehending Score and Retelling	r=.22	94
Beebe	1980	1) Miscue Comprehending Score and Retelling	r=.386	46
		2) Miscue Comprehending Score and Standardized Test	r=.383	46
Sadoski	1981	1) Miscue Comprehending Score and Retelling	r=.620	48
		2) Miscue Comprehending Score and Standardized Test	r=.369	48
Carey	1979	Post-Oral Cloze and Multiple-Choice Questions Standardized Test	r=.66	100
Rankin and Culhane	1979	Cloze and Multiple-Choice Questions (Standardized Test)	r=.68	105
Sadoski	1981	Miscue Comprehending Score and Standardized Test	r=.369	48

TABLE 1 (continued)
Relationships among the Results of Multiple Data Sources of Comprehension

Author	Date	Assessments Utilized	Relationship(s) Reported	N
Ransom	1968	Cloze Test and Informal Inventory at		178
		independent level	r=.4982	
		instructional level	r=.8355	
		frustration level	r=.8112	
W.L. Taylor	1947	Cloze Test and Multiple-Choice Comprehension Scores Over the Same Material	r=.76	--
J.R. Bormuth	1962	Cloze and Multiple Choice Comprehension Scores Over the Same Material at a 4th Grade Level of Difficulty	ranges from r=.77 to r=.99	--

"appropriate cognitive strategies (form-schema) for applying syntactic knowledge to written material" (1982, p. 354), cloze strategies can be especially useful in helping them "recognize and relate subordinating elements . . . to the main idea of the sentence" (1982, p. 354). In addition, the technique forces students to "interact not only with the content words but with function words, the words that establish interconnections among ideas in text" (Seidenberg, 1982, p. 354). A number of modifications on the cloze strategy (zip, maze, synonym, matching) have been designed to encourage just this kind of interaction between the reader and the syntactic constraints of written language.

In brief, reading comprehension appears to involve some level of syntactic awareness. Elkins, Andrews, Apelt, Cochrane, and Atkinson (1979) confirmed this in their report that the grammatic closure subtest of the ITPA (Illinois Test of Psycholinguistic Ability) and another sentence repetition test, both involving syntactic processing, were the only two of a variety of measures they studied which showed a high association with reading ability as measured by a separate reading achievement battery.

While the cloze procedure has been reported to be a useful measure of readers syntactic abilities as well as a useful device for training them to use context, there are numerous indications that it may still "have more to do with the effect of linguistic structures on a reader's

performance than with a generalized measure of reading comprehension" (Heerman & Treadway (1978)).

Miscue. Describing a language-based view of reading, K. Goodman (1970) and F. Smith (1971) helped articulate the importance of examining the "processing of all information available to readers as they attempt to extract meaning from printed material" (Beebe, 1980, p. 326). Comprehending, in Goodman's (1976) view, is a "psycholinguistic guessing game." It involves the reader in predicting meaning based on information that is selected from the linguistic cues of text (graphophonic, syntactic, semantic) as well as in relating text meaning to one's background and experience. Information processing in this view of reading occurs in a top-down fashion. In other words, the construction of meaning does not have to initiate in the reader's use of graphophonic cues alone. Instead, the process starts with whatever knowledge and experience the reader brings to the text and then grows in a flexible and selective use of both linguistic and non-linguistic cues. In this view of reading, getting the author's meaning is the main purpose for interacting with text (Jones, 1982; Kamil, 1978).

Examining the sources of information that a reader may use during reading is possible through a procedure outlined in the Reading Miscue Inventory (Goodman & Burke, 1972). This procedure is a system for analyzing the extent to which errors produced during oral reading are consistent with text

meaning. These errors, most frequently observed in word substitutions are called miscues, include "any incorrect word, partial word, or nonword that is given in place of the correct word in the text" (Beebe, 1980, p. 330).

A comprehending score results from the qualitative and quantitative analysis of oral reading miscues. It is a composite of the percent of all semantically acceptable miscues added to the percent of all semantically unacceptable miscues that are subsequently corrected by the reader. Goodman and Burke (1973) write, "it is assumed that this [system for analyzing miscues] provides insight into the reader's on-going concern for meaning and his success in producing meaningful structures" (p. 32). Both Goodman and Burke (1973) and Carey (1979) conclude, however, that the single best indicator of reading proficiency is the percentage of semantically acceptable miscues a reader produces before any corrections are made. For specific examples of miscues which illustrate readers' use of the grammatical, sound-symbol, and semantic systems of language, see Goodman and Goodman (1977).

Much of the literature on miscue analysis examines apparent differences in the miscueing and correcting behaviors of good and poor readers. In an extensive review of research documenting differences in the comprehension of good and poor readers, Golinkoff (1975-76) reported that poor comprehenders not only produce more errors than good

comprehenders but that their errors are more likely to change text meaning. Likewise, when good comprehenders make errors which do interfere with meaning, they correct them much more frequently than poorer comprehenders do.

Beebe (1980) reported that the best readers in her sample of 46 fourth grade boys corrected their unacceptable miscues almost twice as frequently as poorer readers did. She also observed that all readers tended to refrain from correcting when the miscues produced were semantically acceptable. Carey (1979) concluded that non-correction of meaningful miscues was positive evidence of good comprehending ability. Despite differences in good and poor readers' correcting behaviors, both Carey and Beebe reported that reader failure to correct semantically unacceptable miscues was the single most powerful contributor to poor comprehension as measured by retelling ability and by standardized comprehension test scores.

Miscue analysis has also been applied in analyzing the oral reading errors and correction behaviors of identified learning disabled students. These studies differ from the good reader/poor reader studies because analysis is based on the reading performance of a handicapped population, and because comparisons to non-handicapped populations are not always provided. In Mitchell's (1976) analysis of the oral reading miscues produced by learning disabled subjects, the data led her to conclude that these students' learning

problems do not prevent their use of the grapho-phonic, syntactic, and semantic cues of language when they read. Miscue analysis showed that these subjects tended to use graphic information in the initial portion of words most frequently, followed by the use of graphic information at the end of words. Likewise, LD students' miscues were the same grammatical function as the text word for 71 percent of the 250 miscues analyzed, and considerably more than one-half of the total number of miscues resulted in grammatical structures that were acceptable in prior and preceding contexts. In addition, approximately two-thirds of the total number of miscues produced were semantically acceptable in the entire sentence, while a change in the author's intended meaning was observed in 59 percent of the total miscues produced even though those errors retained the overall syntactic acceptability of the language.

Mitchell's conclusions were based on data collected from a small sample of only ten subjects. Nevertheless, her observation of the correcting behavior demonstrated by learning disabled subjects was consistent with that displayed by poor readers (not identified-handicapped) in other studies. Essentially, neither LD students nor regular students called "poor readers" engage in a significant amount of correcting behavior. This finding corroborates the findings of recent research in the field of learning disabilities (Torgesen, 1977, 1980; Wong, 1979, 1980) which suggests

that LD students tend to be passive, inactive learners who are not involved in monitoring their own learning. In reading, this is especially apparent in LD students' apparent inability to keep track of when they are losing meaning.

In one study comparing miscues produced by 36 LD and 40 non-LD elementary school students, Pflaum (1980) could not identify any oral reading behaviors which detracted from the comprehension demonstrated by non-LD subjects in a retelling of recalled information. With the LD subjects, however, a greater number of miscues resulting in meaning change tended to predict poor comprehension, again as measured by retelling. This finding is consistent with Golinkoff's (1976) conclusions. Pflaum also found that the high phonic cue use demonstrated by LD subjects was a predictor of successful comprehension. Her conclusion tends to substantiate the widely held assumption that the production of fewer decoding errors while reading is generally related to better comprehension. For a more detailed treatment of the relationship between decoding and comprehension behaviors, see Golinkoff (1975-76, pp. 633-638).

Like the other comprehension assessments described in this review, the application of miscue analysis procedures in research settings and in practice is problematic. Three concerns are usually mentioned in conjunction with the use of miscue analysis procedures, including scoring difficulties, dependency on oral production as the major mode of

reading during assessment, and the impracticality of miscue analysis procedures for classroom application.

Specifically, Hood (1975-1976) and Mitchell (1976) attributed scoring difficulties to the existence of different systems for counting and classifying miscues. This makes comparison of the results between different studies nearly impossible. Hood (1975-1976) reported that even when two evaluators were trained to use the same scoring system, the reliability of their decisions about how individual miscues should be categorized was relatively low. Reliability increased, however, when total error scores were computed as opposed to computing errors for different categories of miscues.

The extent to which the demands of oral reading adversely affect the viability of miscue analysis is widely addressed in the literature, particularly with regard to the notion of whether miscues produced during oral reading can be considered representative of the reading behaviors that occur during silent reading (Newman, 1978). Goodman and Goodman (1977), while confirming their belief that the same process underlies both oral and silent reading, write that a reader's desire for accuracy and for speed in the oral reading situation may cause him or her to take fewer chances with unfamiliar words, to change text production in order to cover up miscues, to either limit correction behavior or to over-correct in the interest of word-for-word accuracy, or

to forget completely about maintaining an expectation and a search for meaning (p. 327).

Finally, the practicality of miscue analysis procedures for measuring comprehension in the classroom may be limited by the length of time required to administer, score, and interpret the assessment.

In summary, miscue analysis is generally accepted as a strategy for examining "the influence of language on the processing of visual symbols" (Mitchell, 1976, p. 3), despite some notable limitations. Generally, poorer readers and identified learning disabled students make a greater number of miscues which disturb the flow of meaning and make fewer corrections, overall. Good readers, in contrast, correct many of the meaning-distorting errors they produce and are able to live with those miscues that keep meaning intact. Despite differences between these two categories of readers, two studies reported that identified learning disabled subjects were able to use the cueing systems of language just as their non-handicapped peers did, but that their correction behavior was severely limited, suggesting that their expectation for meaning might be limited during reading and also that they were generally less involved in the process of understanding text.

Patterns of Association among Multiple Data Sources of Comprehension

Introduction. The literature contains numerous references to observed relationships between the results of multiple data sources in reading comprehension (see Table 1). Many of these research findings are summarized in this section, organized according to the possible patterns of association among comprehension measures proposed as three models of comprehension assessment in this study. In discussing the three possible patterns of association among measures, the view of the reading process most closely reflected by each pattern is mentioned, followed by a review of the research which most closely illustrates each one.

Assessment model one: high correlations among the results of four measures of comprehension (questioning, retelling, cloze, miscue). Model one predicts that the nature of relationships observed among four alternative measures of reading comprehension will be at $r \geq .70$. Relationships of this magnitude would suggest that (a) the alternative measures represent "different samples of a single basic skill" (Guthrie, 1973, p. 17), or (b) the alternative measures assess subskills of comprehension which are either identical or "are mediated by a common factor" (Guthrie, 1973, p. 17).

In 1917, E.L. Thorndike suggested that the single basic skill of comprehension was verbal reasoning ability. He

felt that development of this ability was preferable to teaching a number of separate reading skills. Strenecky, Edge, and Strenecky's (1978) review of the factor analytic literature on comprehension confirmed Thorndike's assertion, but concluded that the research also identifies certain linguistic abilities, such as recognizing word meaning and manipulating language, as important components of the process (p. 45). Samuels (1976) also concluded that reading is not a single skill, but may appear to be that way because fluent readers so fully integrate the various subskills of the process that it seems to be a unitary behavior.

Deno, Mirkin, and Chiang (1982) reported the results of a series of three studies designed to identify valid indicators of reading performance which could be easily administered by special education teachers required to make frequent evaluations of student progress toward specific goals. Although they were not looking solely at measures of comprehension, the relationships observed between assessments examined in their studies were high, suggesting that similar components of reading behavior were being tapped by each measure.

In a sample of 18 regular class and 15 LD class students in grades one to five, these authors' correlations of the results of various measures of reading performance showed a high relationship between three oral reading measures (reading word lists, reading outloud only the

underlined words presented in context, and reading a 300 word passage outloud) and two standardized comprehension measures, ranging from r = .73 to r = .91. Correlations were also in the moderate-to-high range for two comprehension measures (cloze, silent reading of 300 word passage plus oral explanation of underlined word meaning) and standardized measures, ranging from r = .60 to r = .83. From these findings, Deno, Mirkin and Chiang concluded that "a simple measure of reading aloud would be a valid index of a student's reading proficiency" (p. 40) and could be easily administered by teachers at frequent intervals.

In the second study in the series, the same researchers examined a new sample of LD and regular class students (N=45) to determine if the observed relationships would change when stimulus materials were varied in difficulty and when different time constraints for test completion were placed on subjects. Results consistent with those of the earlier study were obtained. The researchers concluded that oral reading measures were valid indicators of reading ability and that one-minute samples of reading behavior were sufficient to provide adequate assessment of student progress. In addition they recommended graded oral reading measures as desirable alternatives to the standardized test for "estimating reading level" (p. 44).

The final study in the series was conducted with yet another sample of 66 LD and regular class students to

replicate the earlier findings. Only Words in Isolation, Oral Reading, and Cloze measures were administered to subjects. The results of these measures were then compared to the results of a standardized phonics test and also to the results of two standardized comprehension assessments. While significant positive correlations were observed a third time, the oral reading measure was found to be more highly correlated with both literal and inferential comprehension on the standardized test ($r = .78$, $r = .80$, respectively) than were relationships among the cloze measure and the two standardized comprehension assessments ($r = .67$, $r = .71$, respectively). Table 1 summarizes some of the intercorrelations obtained for studies II and III in the series.

The relationships reported in this series of studies are higher than those reported in any other study found in preparing this review of literature. The researchers did not equate the high associations observed among measures with a view of the reading process that described comprehension in any particular way, however. In fact, the assessment alternatives examined were not obviously drawn from any theoretical description of reading.

Notwithstanding the Deno, Mirkin, and Chiang (1982) conclusions, the view of comprehension as a single basic skill or unitary behavior is not generally supported in the literature as evidenced by high relationships among measures.

Assessment model two: low correlation among the results of four measures of comprehension (questioning, retelling, cloze, miscue). Model two proposes that the nature of relationships observed among alternative measures of reading comprehension will be at $r \leq .30$. Relationships of this magnitude would lend support to a subskills view of reading.

The persistent interest in identifying subskills of reading probably stems from the practical need of researchers and teachers to have a clear focus for assessment and instruction (Otto, 1977; Johnston, 1983). Traditionally used as a framework for structuring remedial reading help, a skills view of reading emerged as one of the earliest conceptualizations of the comprehension process. It continues to be widely used in schools today.

Richek (1978) wrote, "This approach presumes that reading can be subdivided into component skills, that these skills can be inventoried, and that provisions can be made to develop these skills through training" (p. 197).

Farr (1982) indicated that "subskill proliferation" was particularly evident in the assessment of comprehension from the 1940's to the 1960's. During this period, test makers advanced lists of subskills as the basis for measuring students' reading abilities (p. 7). These lists have appeared and continue to appear in the instructional sequences of basal readers and, as such, still have widespread

application in the classroom. Finally, the increased use of criterion-referenced testing and task analysis procedures (in special education particularly) is based on the notion that a behavior can be separated into its component parts followed by the systematic measurement and development of each part to the point that overall competency in a given area will eventually result.

Subskill categorization has not generally contributed to an increased understanding of the comprehension process. On the contrary, confusing terminology among various subskill lists, differences in the sequencing of skills, and different standards for mastery of subskills has fostered divergent views among reading experts about what aspects of comprehension (Farr, 1982) should be measured.

The application of factor analysis referred to in the discussion of assessment model one, has been the primary method for identifying specific components of comprehension on an empirical basis. Many of these components have subsequently appeared on subskill lists. Factor analytic research has a long history, perhaps best exemplified by the frequently reported work of Davis (1944, 1968, 1972).

In his 1968 analysis, Davis provided empirical evidence for identifying the following five behaviors as uniquely distinguishable aspects of comprehension: (a) recalling word meanings, (b) finding answers to questions answered explicitly or in paraphrase, (c) drawing inferences from

context, (d) recognizing a writer's purpose, attitude, tone and mood, and (e) following the structure of a passage. Additional support for the existence of discrete skills has not been forthcoming. For a further discussion of this issue, see Rosenshine (1980) and Spearitt (1972).

Models of reading described by Gough (1976) and LaBerge and Samuels (1976), based on the assumption that the reading process can be separated into specific components or subskills, offered a theoretical basis for the assumption that mastery of word recognition or decoding skills in a certain order (from smaller to larger units of language) is a prerequisite to effective comprehension. Information processing in this view of reading occurs in a bottom-up direction and meaning, in fact, is not important in the process until the decoding of words becomes automatic.

Teachers' uncertainty about whether the mastery of decoding skills must occur before comprehension can occur is apparent in the traditional division between the "code emphasis" and "meaning emphasis" approaches to reading instruction, perhaps best described by Chall (1967). In essence, subskill models of reading continue to be the basis for many teachers' conceptualizations of comprehension, even though they reflect no clear theoretical basis (Farr, 1982, p. 8) and research has provided no evidence that the skills, once identified, are in fact hierarchical (Davis, 1972). Johnson (1983) suggests that there may be a place for the

subskills of comprehension, but only under certain conditions:

> It is clear that it would be useful to know of a set of remediable subskills which comprise reading comprehension. This would provide a framework for effective assessment and remediation of reading comprehension difficulties. If we are to locate such a set we must conduct a search that is driven by theory rather than by solely pragmatic concerns. Thus we must begin to elaborate a coherent theoretical model of reading comprehension (p. 5).

No research was identified in this literature review which reported unusually low associations among measures of reading comprehension. Goodman and Burke (1973) reported a low correlation between the comprehending score derived from miscue analysis and retelling ($r = .22$, $p < .008$, $N=94$). Guthrie (1973), in a study of disabled and normal readers, concluded that while various components of the reading process (from sampling the visual cues of print to integrating text and reader meaning) are qualitatively different, they are nevertheless "interdependent" (p. 11). Based on an analysis of the results of the subtests of a sound-symbol association test, Guthrie found that the intercorrelations between subskills (described as various subtests) were high for normal readers. The same intercorrelations were not

observed for disabled readers, however, particularly for production (versus recognition) tasks (see Table 1). While his generalizations were not based on relationships observed among alternative comprehension assessments, they were considered relevant to this review because of their applicability to the discussion of assessment model II.

Guthrie believed that a reading disability occurred because subskills of the reading process are not mastered to the extent that they can then be fully integrated into a single behavior. The situation of the disabled reader is described by Guthrie in this way, "In this case the individual skills might be acquired at a normal rate for a brief period of time. However, the absence of positive transfer and interfacilitation among subskills would prevent any of the skills from increasing to a normal level of strength" (1973, p. 11). While Guthrie seems to be arguing for the existence of subskills in reading, he nevertheless also argues that, in the last analysis, subskills of the process must be completely interrelated in order for reading to be efficient and effective.

Finally, reading experts have written that there are important subskills of the reading process which have not yet been fully evaluated with respect to their impact on comprehension. Johnston (1983), for example, writes that inferencing, which on Davis' subskill list appeared to be a single, identifiable behavior, is now being shown to have a

number of "fairly well-differentiated types . . . upon which virtually all comprehension is predicted" (p. 6). Kingore and Kurth (1981) also suggest that the cognitive view of reading is now providing a basis for identifying theoretically based subskills of the comprehension process incorporating both bottom-up and top-down information processing models.

In summary, the literature provides some evidence for low association among measures of reading comprehension. Furthermore, while a cognitive theoretical view of the reading process is identifying new subskills of comprehension, they have not yet been reliably measured (Farr, 1982).

Assessment model three: high correlation between the results of two product measures for comprehension (questioning, retelling)/high correlation between the results of two process measures of comprehension (cloze, miscue)/low correlation across product and process assessment types. Comprehension assessment model three proposes that higher interrelationships will be observed within categories of assessment (product and process) then across both assessment categories. Basically, the model suggests that comprehension assessment must involve the evaluation of both the products and the processes of reading because text-based, language-based, and knowledge-based information interact within the reader in the process of comprehending (Rumelhart, 1976). Experts on

comprehension (Harste & Burke, 1978; Hollander, 1975; Royer & Cunningham, 1978; Rumelhart, 1976) conclude that both the products and processes of comprehension are important components of the reading process. Also, they strongly support an assessment of comprehension which allows the interaction of both aspects to be observed.

Theoretically, a high association between measures within product and process categories of assessment but not across both categories would suggest that a different picture of a reader's comprehension behavior may be presented depending on the assessment procedure utilized. This section reviews literature relating to each aspect of the proposed pattern of relationship, including higher correlations between the results of two product measures (questioning, retelling) and between the results of two process measures (cloze, miscue) than are observed across both assessment categories.

Relationships between product measures. The research that follows cites measures of relationship between questioning and retelling comprehension assessments, both product measures. For the studies reported, relationships were positive and ranged from low to high levels of magnitude. All are summarized in Table 1.

Hansen (1978) used a propositional analysis technique to compare the retellings of 17 identified learning disabled fifth and sixth graders with the retellings of 17 regular

education students. Even though subjects were of similar ages, their grade level scores on standardized reading achievement tests differed by approximately two years (LD=$\bar{x}$ 3.77; Regular=$\bar{x}$ 5.96).

After reading and retelling two passages written at a third and at a fifth grade level of difficulty, students were asked a series of comprehension questions which Hansen described as "open-ended and largely factual" (p. 64). In answering the questions, both groups were comparable on the percentage of correctly answered questions for the third grade passage (LD=71.4 percent correct; Regular=76.8 percent correct). On the fifth grade passage, however, regular education students produced much higher scores based on the percentage of correct responses to questions than did the special education (LD) students.

Hansen found significant relationships between number of propositions recalled and percentage of questions answered correctly for the third and fifth grade passages for each group (LD grade 3, r=.61, grade 5, r=.46/Regular grade 3, r=.77; grade 5, r=.50). On the basis of the magnitude of these relationships, Hansen concluded, "proposition analysis appears to be a viable alternative to questioning. The high correlation between the two measures suggests that verbose story retellers tend to score high on comprehension questions" (p. 68). Hansen preferred retelling to questioning because it facilitates a "qualitative analysis of a

student's ability to organize and integrate information" (p. 68).

In a second study reporting relationships between questioning and retelling measures, Adams, Carine, and Gersten (1982) examined the effectiveness of a "systematic study skills training" approach in social studies. The study was conducted by comparing the retelling and question answering abilities of students who had received the systematic study skills training with the similar responses of students who had received either (a) different training, and (b) no training at all.

Forty-five students were randomly assigned to each of three training groups which differed from each other according to the methodology of study skills training utilized. At the end of a specified four-day training period, the effectiveness of the systematic training approach was evaluated by giving all 45 subjects an 800-word social studies passage to study. No constraints were placed on amount of study time. It was hypothesized that the systematic training group would recall more content information in their answers to questions and in their retellings than would the two comparison groups. Subjects were tested on recall of another 800-word passage two weeks later to determine whether the application of study strategies was maintained over time.

Results showed that subjects in the systematic training group performed significantly better on answering questions (10-item short answer test) than the two comparison groups did for both the immediate and delayed measurement situation. On the retelling assessment, the systematic instruction group also did somewhat better than the comparison groups for both the immediate and delayed retellings, but the differences were not statistically significant. The researchers believed that this lack of significance was due, in part, to complications resulting from trying to interpret retellings based on an extremely long passage (800 words). Also, scoring problems resulted in considerable variability in subject's retelling scores.

Regarding associations between measures, a significant, positive relationship was found between the retelling and the questioning assessments in both the immediate and delayed situation (Immediate $r=.32$, $p< .05$; Delayed $r=.57$, $p< .01$). On the basis of these findings, the authors endorsed use of a structured rather than an open approach as a way of facilitating the scoring of retell protocols when using retelling to assess comprehension in subsequent studies.

Relationships between process measures. The research that follows cites measures of relationship between cloze and miscue comprehension assessments, both process measures. For the studies reported, relationships were positive and in the low-moderate level of magnitude. A

summary of findings is contained in Table 1. The comprehending score, originally described by Goodman and Burke in the Reading Miscue Inventory (1972) is generated by adding the total percentage of all semantically acceptable miscues to the total percentage of all semantically unacceptable miscues that are subsequently corrected. Carey (1979) concluded that both of these miscue types were representative of a reader's "successful search for meaning" (p. 12) and suggested that they were also significant indicators of a reader's comprehension ability.

The post-oral cloze test (POC) is administered immediately following the oral reading of a passage and is highly related to the results of conventional cloze tests (Page, 1975). Because of this association, the POC has been used as an alternative to the conventional cloze procedure and is described as a less frustrating assessment task because students read a passage first, and then complete the cloze exercise.

Observing the association between miscue and post-oral cloze measures, Carey (1978) reported relationships in the moderate positive range between the two measures ($r=.55$, $p<.005$, $N=100$). Page (1977) also found that post oral-reading cloze test scores and comprehending scores derived from miscue analysis were positively and significantly correlated ($r=.434$, $p<.01$, $N=48$).

Finally, Sadoski (1980) designed and applied a new scoring system for evaluating oral reading miscues by examining various levels of their semantic acceptability in conjunction with correction behavior, both already demonstrated to be powerful indicators of a reader's comprehension. He reported a positive significant correlation of $r=.547$ ($p < .001$, $N=48$) between a post oral-reading cloze test score and what he called a "comprehension process score" which measured the reader's "semantic monitoring of a given selection" (p. 89).

Relationships across product and process assessment categories. Finally, assessment model three proposes lower relationships among the results of all four measures viewed across both categories of assessment than are observed within the product and within the process categories. The preceding sections summarized the within-category measures of relationship reported in the literature. They ranged from the low-to-high levels of magnitude. A number of studies have also compared the results of various comprehension measures across both assessment categories. An analysis of their findings, summarized in Table I, shows a somewhat mixed picture of the strength of relationships reported.

For example, in associating the results of "process" comprehending scores (derived from miscue analysis) and "product" retelling scores, two studies reported low positive correlations of $r=.22$, $p < .008$, $N=94$ (Goodman & Burke,

1973) and r=.386, p < .003, N=46 (Beebe, 1980). Another study (Sadoski, 1981), relating the results of similar measures, reported a moderate positive correlation, r=.620, p < .001, N=48.

Two studies (Carey, 1979; Rankin & Culhane, 1979) indicated a moderate positive correlation between the results of a cloze and a questions assessment (defined as a multiple choice test or "product" measure). Rankin and Culhane reported an average r of .68, p< .01, based on the responses of a sample of 105 fifth graders. While Carey's results were based on the computation of a relationship between a post-oral cloze measure and a standardized test, the correlation was of a similar magnitude to Rankin and Culhane's (r=.66, p< .005, N=100).

In a comparison of readers' comprehension as measured by both cloze and retelling assessments, Rigg and Taylor (1979), in a study discussed earlier in this review, observed that as subjects' retelling scores went up their cloze scores also went up. The number of retellings produced (none, one, or two) and the timing of those retellings (immediate, delayed) were not associated with gains in cloze scores. While the authors did not report specific measures of relationship between these two measures, they pointed out that retelling seemed to "require a 'deeper' cognitive processing than that necessary to . . . fill in cloze blanks" (p. 82).

Finally, a low positive correlation ($r=.369$, $p < .005$) was reported by Sadoski (1981) in a correlational analysis of miscue scores and a specially constructed passage-dependent multiple-choice questions test, similar to a standardized test, on a sample of 48 fifth-grade readers. Beebe (1980) reported nearly identical results in correlating miscue scores and standardized test scores ($r=.383$) in a sample of 46 fourth grade boys.

Summary. This review of relationships among assessments did not unequivocally confirm a high association ($\geq .70$) or a low association ($\leq .30$) among multiple measures of reading comprehension. Essentially, no consistent pattern of association among measures was evident, yet there was some evidence in the literature supporting each of the possible patterns of association hypothesized in this study.

A recapitulation of research findings (see Table 1) shows that the relationships between two measures of comprehension identified as product assessments (questioning, retelling) may be somewhat higher than those observed between the two measures of comprehension identified as process measures (cloze, miscue), although no clear-cut pattern was observable. Differences among the various studies with regard to variables such as directions given to subjects, timing of assessments (immediate, delayed), level of difficulty of materials read, instruction given prior to assessment, and variation in subject ability, among others,

makes direct comparison of results difficult. These differences may account, in part, for the lack of a consistent pattern in the findings among studies.

Further, with regard to relationships observed across product and process categories of assessment, lower associations between miscue and retelling scores were observed in comparison to those observed between miscue and the cloze scores. No relationships between miscue and questioning measures were reported in the literature, unless one considers the standardized test a substitute for the questioning assessment. In that case, miscue scores are related to standardized test scores at a low-positive level. Cloze, on the other hand, is moderately associated with the standardized test, suggesting more linkage between cloze and questioning then is observed between miscue and questioning.

A major conclusion of this portion of the literature review is that no one study reported has used the same subjects across five different measures of comprehension, has classified measures as product and process assessments, or has proposed possible patterns of association among measures. In view of the inconclusive findings in the literature, the present study was needed to investigate relationships among a wider range of multiple measures of comprehension, to articulate models of assessment that represented possible patterns of associations among those measures, and to test for the existence of specific patterns

of association. The fact that subjects in the present study had been identified learning disabled added another dimension to the examination of relationships among alternative measures.

Learning Disability as a Handicapping Condition

Definition and characteristics. In 1975, Public Law 94-142, the Education for All Handicapped Act, was passed by the United States Congress. Insuring the right to a free appropriate public education for all handicapped individuals from ages 2 to 21, the law provides a definition of various handicapping conditions that are to be used by state and local education agencies as guidelines for identification. The definition of learning disability that appears in the law was accepted by the State of Virginia. It reads:

> Specific learning disability means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations which adversely affects educational performance. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the

> result of visual, hearing, or motor handicaps, or mental retardation, or emotional disturbance, or of environmental, cultural, or economic disadvantage. (PL 94-142, 121a.5)

Even with the help of this definition, the problem of distinguishing students who have a "learning disability" as defined in Public Law 94-142 from those who exhibit some other problem that affects school achievement has persisted (Artley, 1980; Gredler, 1978; Harris, 1980; Kaluger & Kolson, 1978; Richek, 1978). In addition, since underachievement in reading is the main reason to suspect that students might have a specific disability (Gredler, 1978), deciding who has a reading problem and who has a learning disability that manifests itself in a reading problem has been especially difficult.

Kaluger and Kolson (1978) suggested that "The question whether there is a difference between reading disabilities and learning disabilities theoretically could be resolved by adequately defining each field. But that, in part, is just where the problem lies--there is no truly adequate definition of learning disabilities" (p. 2) even though the legal definition provides some direction. Even the legal definition, according to these authors, "falls short of specifically pinpointing the learning disabled child" (p. 2).

Harris (1980) reported that the preliminary drafts of Public Law 94-142 had included specific guidelines for

comparing a student's actual achievement with estimated potential. According to these guidelines, a "severe discrepancy" in achievement would be considered anything less than 50 percent of a student's estimated potential (Harris, 1980, p. 421). These guidelines were never incorporated into the final legislation, however, possibly because the characteristics of learning disabled individuals were far too complex to be adequately accounted for and explained by a mathematical formula. The current result of such a limitation is that the criteria for determining that students are learning disabled are frequently vague and extremely hard to document conclusively through available psychoeducational assessment devices.

Kaluger and Kolson (1978) further concluded that most definitions of learning disability typically agree on four major aspects: "(1) Intellectual capacity is average or better; (2) There is a discrepancy between expected and actual achievement; (3) There is a disorder in one or more of the basic psychological processes involved in using spoken or written language; and (4) The deficiencies are not primarily due to visual, hearing, or motor handicaps; to mental retardation; to emotional disturbance; or to environmental disadvantage" (p. 3). In addition to these five aspects, learning disabled students may also have a notable birth or developmental history, may exhibit certain behavioral characteristics such as hyperactivity,

hypoactivity, distractibility, and inconsistency of performance, may have physiological involvement such as body laterality problems and poor coordination for fine and gross motor movements, may have speech and language development disorders, and may have difficulties with basic thinking processes, among other problems.

Bryan (1974) suggested in a review of the research comparing LD and non-LD learners that simple perception and discrimination, hyperactivity, and neurological deficits did not discriminate between these two groups of students. Instead, she concluded that the research showed the characteristics which distinguished best between groups were attending ability, complex tasks with a heavy language component, and ability to organize information. For a further discussion of the characteristics of learning disabled students, particularly as they relate to the development of reading competencies, see Gillespie-Silver (1979).

Trends in the field. The field of learning disabilities had a medical orientation at its inception (Richek, 1979) which was the result of an early interest in investigating both the brain functioning of children as well as the results of traumatic injury as possible explanations for poor academic performance. Development and application of the term and concept "minimal brain damage" resulted from these early investigations. This concept was further expanded upon and subsequently associated with the terms

"learning disorder" and "learning disability," and finally was related to the education context in Kephart's The Slow Learner in the Classroom (1971).

The major assessment/intervention strategy used with LD learners with reading problems through the years has reflected this medical orientation of the field. Called the process training approach, it focused primarily on developing an individual's basic abilities in the perceptual, sensorimotor, and language areas. These abilities were believed to be important prerequisites to the actual teaching of reading and were sometimes taught to the exclusion of any academic instruction. As Zigmond, Vallecorsa, and Leinhardt (1982) report, "Advocates of this approach hold that the disabled reader lacks certain prerequisite abilities but does not lack the ability to learn to read" (p. 8). Process training has not been widely supported through empirical studies in the literature (Balow, 1971; Hammill & Larsen, 1974).

Other approaches to teaching learning disabled students required the teacher to determine whether the student was an auditory or a visual learner and then to match reading instruction to the strongest channel for learning. Usually this approach meant that so-called auditory learners would be taught phonics strategics for word-recognition where visual learners would be taught through a whole word (sight) method.

Yet another approach to instructional intervention in reading was the application of whatever approach was considered to be the "best method" (Zigmond, Vallecorsa, & Leinhardt, 1982, p. 94) for teaching identified learning disabled students. These "best methods" included the widely known approaches of Fernald and Orton-Gillingham as well as those represented by the Distar program, the language experience approach, as well as neurological impress, individualized reading, and alternative alphabet approaches. Despite this abundance of approaches to instruction, Zigmond, Vallecorsa, and Leinhardt summarized that "little is known about whether systematic variation in the materials, content, sequences, or strategies of reading programs produces different learning outcomes in different students" (p. 92).

In recent years there has been a gradual shift away from the process training emphasis in teaching LD youngsters. Harris (1980) describes a shift "in the past few years from training of basic abilities such as visual and auditory perception to direct teaching of academic skills and knowledge" (p. 423). An example of this shift is illustrated in the content of the literature itself with journal articles prior to 1970 being primarily concerned about research in the perceptual-motor areas and how those findings were related to learning. This is not the case today, however, as reflected in a review of literature by

Torgusen and Dice (1980) which illustrates movement toward a greater consideration for aspects of memory and of the cognitive functioning of the learner that can be directly related to any academic area, but especially to reading (pp. 531-532). Chall's (1978) review of ten years of research in reading and learning disabilities traced the same trend.

In short, students who are disabled readers can find themselves in one of two places in school. They can be identified as "learning disabled" (with somewhat vague criteria) in which case they may still be engaged in more process training activities than reading activities in their special education classes. Or, they may remain unidentified but receive special assistance from the reading teacher to remediate reading problems. Whichever happens, problems created by the "overlapping jurisdictions" (Harris, 1980, p. 422) of the learning disabilities teacher and the reading teacher in addressing the needs of students who have trouble reading has yet to be practically addressed, much less resolved.

Comprehension research with LD readers. Studies of the reading performance of LD students agree that, as a group, these students do not achieve well, particularly as measured by current testing procedures (Ysseldyke, 1983, p. 231). Other studies with LD subjects reported in the questioning, retelling, cloze, and miscue sections of this review confirm this generalization.

A number of studies describing the reading behaviors of learning disabled subjects show that LD students are not proficient language users in comparison to their peers. In a comparison of normal and dyslexic second grade boys' use of syntactic cues in reading, Vogel (1977) found the dyslexic group to be significantly deficient in their ability to inflect real and nonsense words through use of morphological rules of language. This finding suggests that differences between the two groups might be observable at an early age and also that that "poor readers may not be utilizing to the fullest the semantic and syntactic information provided in the morphology of written language" (p. 41).

White, Pascarella, and Pflaum (1981) documented a positive relationship between the use of syntactic functions of language and reading comprehension. Thirty identified LD students were randomly assigned to two syntactic training groups. One group was taught a sentence combining strategy based on angram use and on the active manipulation of words and phrases to construct progressively longer, more meaningful sentences. The second group was taught a variety of sentence patterning activities (matching noun and verb phrases, identifying type of unpunctuated sentence, combining sentences with the conjunction "and", and so forth) more typically a part of regular classroom reading instruction. Age, IQ, and mean comprehension subtest scores on a

standardized test were not widely different between the two groups. Results of data analysis showed that the sentence construction activities of the angram treatment group appeared to significantly affect students' comprehension performance on a post-treatment sentence construction and cloze test in comparison to the traditional strategies group. Gains were particularly noticeable for subjects whose minimum decoding skills were measured at the second grade level of functioning or better. This study showed that "the learning disabled readers learned to manipulate elements within sentences on the basis of sentence meaning" (While, Pascarella, & Pflaum, 1981, p. 703).

In a general sense, Zigmond, Vallecorsa, and Leinhardt (1982) found that LD students' achievement in reading was more highly related to the quality of teachers' instruction and to an increase in "on-task reading time, especially silent reading, by as little as five to 10 minutes each day" (p. 96). These findings seem especially important in light of Van Etten's (1978) observation that where reading skills have been developed in the LD classroom, decoding has been more frequently stressed than has comprehension. A heavy emphasis on decoding in the LD classroom reading program might leave little time for actual reading.

Further, learning disabled students appear to have particular trouble with expressive language in addition to trouble understanding language receptively (Newcomer &

Magee, 1977). The language production of 38 identified elementary age LD students, all of whom had reading problems, was compared to the performance of an equal number of non-identified good readers of the same age. The mean IQ was 102 for the LD group and 104 for the good reader group. Results of students' performance on a test of language performance confirmed the significantly higher language production abilities of the good readers. While none of the non-identified group scored below the mean on any subtest of the language production test, the LD group showed much variability within subtests, having "less difficulty with receptive linguistic skills, which involve pointing responses, than they have with expressive tasks" (Newcomer & Magee, 1977, p. 899). The authors concluded that if a child's oral language use is not proficient in comparison to his peers, he may be more likely to demonstrate reading difficulty.

Blumenthal (1980) studied the naming competence in a timed test situation of 91 learning disabled students in the second through fifth grades as an indicator of verbal retrieval ability. She then related these findings to the demonstration of reading strategy use on the Reading Miscue Inventory (Goodman & Burke, 1972). Although no specific measures of relationship were reported, the author cited the absence of an association between the two variables under study. While the LD students studied did produce many more

retrieval errors than the normal population on which the retrieval test was normed, their oral reading miscues suggested that they "can and do use all available linguistic cues" (Blumenthal, 1980, p. 44). Mitchell (1976) also observed LD students using grapho-phonic, syntactic, and semantic cues of language during oral reading even though they did not display as great a tendency to self-correct as good readers did and their overall reading achievement was comparatively lower.

Finally, Hansen (1978) in a study reported earlier in the chapter, found that LD students answered fewer questions and retold fewer propositions about a passage than non-handicapped students did. While regular education students retold more main ideas than did their LD counterparts, the two groups did not differ in the retelling of details. Neither did they differ on the inclusion of inferred information in retellings. Hansen's findings further confirm the generalization that many LD students with reading problems do demonstrate comprehending abilities, particularly when text material is not too difficult. On the other hand, their overall achievement in reading is not usually strong in comparison to non-identified learners. Inconsistency and variability in performance appear to be distinguishing characteristics of LD students as a group and their competency in oral and in written language is not commensurate with what is displayed by non-identified students.

Summary. While students who are called learning disabled share some characteristics in common, Ysseldyke (1983) expressed the view that these students "are learning disabled only when they are not other things" (p. 226). The limited understanding of the exact parameters of a learning disability is still pervasive in the field of special education.

Despite this continuing confusion over definition and description of the handicapping condition, research in reading using learning disabled subjects has shown that these students' overall performance is lower than non-identified students and that they generally have trouble with language-related tasks, particularly in the areas of manipulating the syntax of language, oral expression, and monitoring their own ability to keep the meaning of text going. On the contrary, research has also demonstrated that learning disabled students do appear to use linguistic cues to generate meaning when reading and can demonstrate their comprehension when readability of text is not too high.

Finally, the trend in instructional programming for LD students is currently embracing direct teaching in specific academic areas rather than training in the basic psychological processes. To that end, it is imperative that continuous, frequent, and direct assessment of student progress toward identified instructional goals be part of any special education program, especially in response to

specifications of the legally required IEP or Individualized Education Program (Deno, Mirkin, Chiang, 1982, p. 36).

Summary of Chapter Two

The potential for flexibility in assessment has contributed to an interest in understanding how comprehension can be measured most effectively. The assessment of comprehension has received attention in the literature because it is an area of inquiry that directly relates to refining conceptualizations about the nature of the reading process itself. Behaviors that are part of the comprehension process have been shown to be too complex and too hard to observe directly to be adequately measured in one standardized test or by asking a series of questions. Hence, the use of alternative indicators of comprehension has been suggested (Page & Vacca, 1979) not only as a diagnostic strategy, but also as a way of learning more about how readers get meaning from text.

The current state of the literature regarding the assessment of comprehension with multiple measures shows that:

1. Comprehension assessment is problematic because the construct of comprehension is complex and has not been consistently defined on the basis of a single theoretical view of the reading process on which most reading experts can agree.

2. Assessing the products of comprehension usually involves measuring the reader's exact reproduction of text meaning. Answering questions and retelling passage meaning are frequently used product indicators of comprehension.

3. Assessing the processes of comprehension usually involves making inferences about how the reader uses knowledge of text, knowledge of language, and knowledge resulting from background/experience information to construct text meaning. Cloze exercises and miscue analyses are frequently used process indicators of comprehension.

4. Movement toward reaching consensus in defining comprehension is being made. This consensus involves agreement among reading experts that linguistic information (text), cognitive/linguistic information (reader), and a range of other information resulting from a reader's prior knowledge and background all affect comprehending ability in an interactive way and should be addressed during comprehension assessment.

5. Multiple measures of comprehension are currently in use which differ in assessment task demands and in the scoring and interpretation of data.

6. The literature documents a range of relationships among multiple measures of comprehension.

7. There is no clear pattern of relationships among multiple measures of comprehension (either high associations among multiple measures, low associations among multiple

measures, or higher assocations among measures within a given category of assessment (product-process) than among measures not grouped by category of assessment).

The current state of the literature regarding the field of learning disabilities shows that:

1. The varied characteristics of learning disabled students complicate a precise descrip tion of the handicapping condition.

2. Learning disabled students generally score lower on measures of reading comprehension in comparision to regular class students.

3. Learning disabled students are typically not good language users; they have trouble manipulating the syntax of language; their expressive language is frequently weak; they tend not to monitor their own reading behaviors, but they can be taught to do these things.

4. There may be developmental differences in the capabilities of good and poor readers to understand text.

5. Learning disabled students do demonstrate the same comprehending abilities as regular class students when the difficulty level of reading materials is not too high.

6. Assessment and intervention strategies used in the LD classroom currently emphasize developing skill in academic areas (especially reading) rather than developing skill in the basic psychological processes (auditory, visual and kinesthetic perception and processing) of learning.

7. Research in learning disabilities has moved toward support for a cognitive view of the reading process.

CHAPTER 3

Procedures

Introduction

In order to investigate the nature of relationships between and among alternative measures of reading comprehension and to evaluate three models of comprehension assessment which hypothesize the specific nature of those relationships, the reading comprehension of seventh and eighth grade learning disabled males was examined.

This chapter will review the procedures involved in carrying out the investigation. First, selection of the sample will be described, followed by a summary of sample characteristics. Next, the procedures for designing and redesigning four different reading passages, for developing alternative comprehension assessments to accompany each passage, and for collecting the data will be reviewed. Finally, the research hypotheses and data collection and analysis plans will be described.

Sample

Selection of the sample. The subjects in this study were students in a school district located in the metropolitan area surrounding Washington, D.C. Subjects were males in either the seventh or eighth grades who were identified as learning disabled according to the school district's criteria which were, in turn, based on State and Federal regulations resulting from Federal Public Law 94-142, The Education for All Handicapped Act of 1975.

Initially, it was intended that students would also meet the following requirements for inclusion in the sample: (a) a minimum reading grade level of 3.0 as measured by a standardized achievement test recorded in the "Current Level of Performance" section of each student's Individualized Education Plan or IEP, and (b) a Full Scale Intelligence Quotient (IQ) score within the average range (90-109) as measured within the past three years by the Wechsler Intelligence Scale for Children (Wechsler, 1974).

The criterion relating to IQ was later modified in the interest of achieving an adequate sample size. While 22 subjects (54 percent of the sample) had Full Scale IQ's in the average range, one student was measured in the mentally deficient range (69 and below), three students in the borderline range (70-79), 12 students in the low average range (80-89), four students in the high average range (110-119), and two students in the superior range (120-129). This change has implications for the generalizability of the findings.

One of the four intermediate schools in the district with seventh and eighth pupil populations did not allow students to be screened for participation in the study. This caused a reduction in the total number of students within the district who were eligible to participate. Hence, random selection of subjects was not possible.

Forty-eight students were finally selected for the sample. An additional student was held in reserve in the event of subject attrition from the original group.

Confidentiality of information was ensured in accordance with the procedural safeguards guaranteed to identified handicapped students by Federal Law (P.L. 94-142). Student numbers were utilized in lieu of names to guarantee complete anonymity of participants.

Characteristics of the Sample

School. Of the three schools which had students participating in the study, the largest number of subjects attended school one (27 or 56 percent).

Grade. A greater proportion of seventh graders (58 percent) than of eighth graders (42 percent) was noted in the sample.

Age. The mean age of subjects in the sample was 174 months or 14.5 years, ranging from a minimum age of 157 months (13 years) to a maximum age of 203 months (17 years). The most frequently occurring age was 169 months or 14 years.

IQ. The distribution of Full Scale WISC (Wechsler, 1974) IQ scores showed the mean IQ for the sample to be 95, within the average range (90-109), with a standard deviation of 12 IQ points. The Wechsler was standardized on a mean of 100 and a standard deviation of 15 IQ points. Over 95 percent of the sample in the study had measured Full Scale IQ's

within ± two standard deviations from the mean of the WISC standardization sample.

Reading grade level. Confidential files at each school were reviewed to verify that each student included in the study had achieved a minimum reading grade level of 3.0 on some objective measure of reading ability other than the standardized measure utilized in the study. For most subjects, the objective measure utilized was either the Woodcock Johnson Psychoeducational Battery (Woodcock & Johnson, 1977) or the Woodcock Reading Mastery Test (Woodcock, 1973). Both of these tests were frequently used by the school district to obtain a current measure of students' educational performance in reading and in mathematics. The current reading level of 77 percent of the sample was within the third-to-fifth reading grade level range. This was consistent with expectations for a sample of learning disabled students who frequently have a history of reading problems (Zigmond, Vallecorsa, and Leinhardt, 1982). A much smaller portion of the sample, twenty-three percent, was found to be reading within the sixth-to-twelfth reading grade level range.

Summary of sample characteristics. In summary, over one-half the subjects in the study attended one school. Seventh graders represented the largest portion of subjects in the sample. Subjects' mean age was 175 months, or roughly 14.5 years. Considering the greater proportion of

seventh graders in the sample, this mean age seems high for an intermediate school sample, and particularly high for seventh graders who were represented more frequently in the sample than were eighth graders. A mean age of 14.5 years was consistent with expectations for a handicapped population, however, because many students had been retained for at least one and sometimes two years during their elementary school years. When viewed as a whole, the mean Full Scale WISC IQ for the sample was in the average range of intellectual functioning. This was not the case for all subjects since 22 of them had measured Full Scale IQ's either lower or higher than the 90-109 score limits defining "average." The reading grade level of most subjects in the study was within the third-to-fifth grade level range.

Development of the Passages

The four narrative passages read by each subject during the course of the study were adapted from the Skillbooster series (Modern Curriculum Press, 1975), the Basic Comprehension Series Breakaway (EMC Corporation, 1980) and Spotlight Scope Reading Skills 3 (Scholastic, 1973). While these comprehension materials were generally available in the school district, they had not been recently used for instruction or as a source of practice activities in the reading classes of the three schools where subjects were enrolled.

The adaptation process was accomplished by first identifying within various comprehension practice materials particular selections that were similar in overall difficulty, length, and interest level of content. The actual text language of the identified selections was then modified somewhat to approximate a passage length of around 350 words and passage readability between the 3.5 and 4.0 level of difficulty as measured by the Fry readability graph (1968). Given the handicapping condition, age, and current level of performance of subjects in the study, it was reasoned that maintaining passage readability at the 3.5-4.0 level provided text that would be challenging enough to generate the number of errors necessary for evaluating differences in comprehension, but not so difficult as to make the reading task overly frustrating. No initial attempt was made to equate number of idea units across each of the passages selected, although this factor later became an important aspect of the scoring procedure for one of the comprehension assessments investigated in this study. Four separate passages were subsequently developed as a result of this process.

The results of a field test confirmed that the readability level of the four reading passages was too low. (Note: These passages will be made available to researchers studying younger, normal, or disabled readers.) Based on this experience, a new set of stimulus materials was

developed. Four new passages were selected from the same comprehension practice materials as the first four passages and were adapted in the same way. Two alternative readability formulas (Flesch, 1948; Gunning, 1952) were applied to determine the difficulty levels of the new passages. New passage readabilities ranged from grade 5.1 to 6.3 in difficulty and their length ranged from 342 to 349 words.

Overall, the new passages were judged satisfactory as stimulus materials for reading comprehension assessments. Table 2 shows the readability level of each passage according to the two formulas specified. In addition, the total number of idea units in each passage and passage length in number of words are shown. Each new passage used in the study can be found in Appendix A.

Alternative Comprehension Assessments

Four ways of measuring comprehension were examined in this study. To determine whether subjects' comprehension varied under alternative assessment conditions, questioning, retelling, cloze, and miscue comprehension assessments were developed for each of the four passages. Consequently, sixteen sets of reading passage-plus-assessment stimulus materials were prepared (i.e., four conditions times four passages). Questioning and retelling tasks were classified as product measures of comprehension while cloze and miscue tasks were considered process measures. Chapters One and Two describe the differences in these two categories of measurement in detail.

TABLE 2

DESCRIPTIVE DATA ON NEW TEST PASSAGES

Test Passages	Flesch Readability Formula	Gunning Fog Readability Formula	Total No. of Idea Units	Total No. of Words (passage length)
"Give the Ball to Wilt" (Passage A)	6.3	5.8	30	347
"Harriet Tubman" (Passage B)	6.3	5.2*	38	347
"First in the Sky" (Passage C)	5.6	5.1	31	349
"Fine Animal Gorilla" (Passage D)	6.3	5.8	34	342

*Note. This index of readability would have been 6.5 had the formula incorporated proper nouns of three or more syllables into the computation.

The following section describes the development and scoring of each of the four comprehension assessments. In addition, a formal standardized measure of reading comprehension used in the study is also briefly discussed.

Assessment with Questioning

Development of questions. One form of assessment of subjects' comprehension was accomplished by asking questions following silent reading. A series of 10 questions was developed for each of the four passages according to a taxonomy of question-answer relationships outlined by Pearson and Johnson (1978) which includes textually explicit, textually implicit, and scriptally implicit questions (p. 157). Answers to textually explicit questions are available in the language of the text itself. To answer textually implicit questions, however, the reader must do more "reading between the lines." "Comprehension is regarded as textually implicit if there is at least one step of logical or pragmatic inferring necessary to get from the question to the response and both question and response are derived from the text", (Pearson & Johnson, 1978, p. 161). Finally, scriptally implicit questions require the reader to call on his/her own knowledge and background of experience, or "script", in order to generate an answer (Pearson & Johnson, 1978, p. 157). Of the 10 questions developed for each of the four passages, four were textually explicit, four were textually implicit, and two were scriptally implicit.

Scoring. An answer key for the 10 questions accompanying each passage was developed. Subjects received one point for each correctly answered question. No partial credit was given. Revision of the answer key was undertaken when the need became evident during testing with all responses then being rescored to reflect these changes. Points were then totaled to achieve a raw score for the questioning assessment condition. Raw scores could range from a minimum of zero points to a maximum of points.

Assessment with Retelling

Development of retelling outlines. Comprehension was also assessed through oral retelling of passage content following silent reading. Four separate content outlines based on the number of idea units contained in each passage were prepared according to discourse analysis procedures described by Taylor (1978). Level one idea units for each passage were classified as main ideas and were recorded on the left side of the content outline for each passage. Level two and level three idea units for each passage were classified as supporting details. These two categories were combined and were recorded on the right side of the content outline for each passage.

To establish reliability in the discourse analysis of passages, the researcher and a second examiner made independent identifications of idea units at levels one, two, and three for each passage at the initial stages of content

outline preparation. Results of the independent analyses were compared. A preliminary agreement on the total number of level one idea units for passages A, B, C, and D was 45 percent, 42 percent, 55 percent and 100 percent, respectively. Following discussion to resolve differences in the application of scoring procedures, the level of agreement between the two analysts on level one units for passages A, B, C, and D was 64 percent, 71 percent, 66 percent, and 100 percent. Level two and three idea units were subsequently identified following agreement on the level one units.

Table 2 shows that the total number of idea units to be retold differed only slightly from passage to passage.

Scoring. The content outlines prepared for each passage were used as scoring guides to determine the extent to which ideas expressed in subjects' unaided, oral retellings matched the passage idea units identified through discourse analysis procedures. As subjects included ideas in their retellings, they were given one point for each main idea or supporting detail retold.

Specific guidelines were followed in evaluating each idea unit retold. Full credit (one point) was given only when (a) subjects retold the subject, verb, and object (if any) in an idea unit, (b) subjects used a verb in retelling an idea unit which expressed the same concept as the verb in the text, and (c) subjects retold a total of two out of three parts of a lengthy idea unit. These guidelines were

adapted from those suggested by Taylor (1978, pp. 108-111). Half-credit was not given for the partial retelling of an idea unit.

No specific arrangements were made for scoring subjects' retelling of information which did not appear in the content outlines. Although the retelling of unanticipated information occurred infrequently, scores on this measure did not reflect such information. All unusual or otherwise interesting inferential statements were recorded verbatim in the margin of the scoring guides. These are discussed in the Supplementary Analysis section of Chapter Four.

A raw score was calculated for the total number of level one idea units retold, for the total number of combined level two and three idea units retold, and for the grand total of idea units retold. This grand total of idea units retold was the only score used for the retelling assessment. Raw score data were then converted into percentages in order to reflect the variation in total number of idea units associated with each of the four passages. Percentages of total idea units retold could range from a minimum of zero percent to a maximum of 100 percent.

Assessment with Cloze

Development of cloze passages. Subjects' comprehension was also assessed through application of the cloze technique. Each of the four reading passages was prepared in a

conventional cloze format. The initial sentence was left intact, followed by an every-tenth-word deletion pattern suggested by Karlin (1973) as more appropriate to informal diagnostic assessment than other deletion patterns. Twenty-five cloze blanks were created for each passage following this deletion system. Beyond the twenty-five blanks, passages reverted to unmutilated text.

Scoring. Answer keys were developed for each cloze version of the four passages, and a raw score based on one point for each exact replacement of omitted words in a given passage was calculated for each subject. This total represented the overall score obtained for the cloze assessment condition and ranged from a minimum of zero points to a maximum of twenty-five points.

Assessment with Miscue Analysis

Development of a miscue analysis system. Miscue analysis of the first 25 miscues produced by each subject during oral reading of a passage was the fourth procedure examined in the study. The first 25 substitution miscues produced by each subject were listed on a worksheet copy of the passage, using a recording system described by Goodman and Burke (1972).

Scoring. A qualitative analysis of miscues was then accomplished by applying specific scoring criteria outlined in procedures for using the Reading Strategies Profile (Hutson & Gove, 1980). Although miscues were analyzed for

graphophonemic and syntactic acceptability, evaluation of their semantic acceptability was the only factor considered in generating each subjects' total score for the assessment.

A specific scoring arrangement was made to provide for numerical comparability among subjects' scores because all of them did not produce the maximum number of miscues possible (25). Each subject's observed score for the miscue assessment was based on two factors: (a) an automatic assignment of three points for each non-miscue, and (b) an assignment of zero, one, or two points for each miscue produced based on a qualitative evaluation of its semantic acceptability. According to Hutson's criteria (Hutson & Gove, 1980), zero points were assigned for semantic acceptability if a miscue did not make sense in the phrase, one point was assigned if a miscue did make sense in the phrase but was not synonomous with text meaning, and two points were assigned if a miscue resulted in a phrase which was nearly synonomous with text meaning. (All miscues were evaluated according to their semantic acceptability within a phrase or clause.)

The total miscue score was computed with the following formula: 3(25 - No. of miscues evaluated) + Semantic Acceptability Score. For example, if Subject A did not miscue at all in his oral reading of a passage, his total score would be 3(25-0)+0=75 points. This score (75 points) was the maximum score possible for the miscue assessment. Further,

if Subject B miscued eight times in his oral reading and the qualitative analysis of those eight miscues resulted in a "Semantic Acceptability Score" of two points, his total score would be 3(25-8)+2=53 points. Finally, if Subject C produced 25 miscues, none of which "made sense in the phrase" (Hutson & Gove, 1980), his total score for the assessment would be zero points or 3(25-25)+0=0 points, the minimum score.

Finally, the semantic acceptability of each miscue which had been self-corrected by the subject was scored again according to the point system described. Each subject's total number of points for the miscue assessment was then recalculated to include all self-corrections produced. While these recalculations produced relatively minor score changes overall, the score including self-corrections was the one used in the major analysis of data, realizing that self-corrections are not always fully correct. The total score for the miscue assessment, taking into consideration all self-corrections, could range from a minimum of zero points to a maximum of 75 points.

Inter-rater Reliability in Scoring

Since evaluating subjects' performance on the retelling and miscue assessments each involved an element of subjectivity, both were scored twice to determine the reliability of scoring procedures. To accomplish this, a random sample of one-third (N=16) of the retellings was scored by a rater

other than the researcher. A second scorer analyzed the miscues produced by another random sample of one-third (N=16) of the subjects. The Pearson Product-Moment coefficient of correlation calculated between the researcher and the scorer for percentage of total idea units retold on the retelling assessment was r = .85. For the semantic acceptability of miscues evaluated by an adaptation of Hutson and Gove's (1980) procedures prior to self-correction, the Pearson r between the researcher's and the second rater's scores was calculated at r = .98.

No second rating of the questioning and cloze assessments was necessary because of the relative objectivity involved in judging the correctness of those responses.

Standardized Comprehension Test

One formal measure of reading comprehension, the Gates MacGinitie Reading Test, Level D, Form 3 (Houghton Mifflin, 1978) was also used in this study. This test was assumed to be representative of most standardized reading achievement tests that are normally used in practice as indicators of reading comprehension. The Gates MacGinitie test was standardized on a norming group of a stratified random sample of approximately 6,000 students who were representative of different geographical locations within the United States, of both urban and rural settings, and of various socioeconomic and ethnic backgrounds. The technical manual accompanying Level D did not base discussion of test validity on

any theoretical description of the reading process. It stated that the test was intended to measure "important knowledge and skills that are common to most school reading curricula" (MacGinitie, 1978, p. 38), but a more precise description of what behaviors were associated with such knowledge and skills was not provided. Kuder-Richardson Formula 20 reliability coefficients for the comprehension subtests of Level D, forms 1 and 2, were both .91. These coefficients were calculated on the standardization sample for Level D. No reliability coefficient was reported for Level D, Form 3 of the test.

The raw score for the comprehension subtest of the Gates MacGinitie Reading Test, Level D, Form 3 (Houghton Mifflin, 1978) was used in the main analysis for this study. While the tasks in Level D are designed for grades 4-6 and therefore are generally inappropriate to seventh and eighth graders, they were judged by the researcher to be more appropriate than the tasks in Level E for measuring the comprehension of subjects in this study. This decision was based on the fact that subjects' measured reading grade levels were below their actual grade placement levels in a majority of cases. In any case, out-of-level testing seemed appropriate since interpretation of subjects' comprehension performance in relation to the norms resulting from the standardization of the Gates MacGinitie Test was unimportant for the study.

Data Collection Procedures

A written outline of the study including its purposes and design was presented to the school district administrator in charge of research for review prior to data collection in the fall of 1981. Following administrative approval, details of the study were described to intermediate school principals and staff cooperation in carrying out the research plan was requested.

Before data collection began in January, 1982, the special education staff in each of the three participating schools met with the researcher to discuss overall project organization. Because subjects worked with more than one teacher at each school site, a mutually convenient time for data collection was negotiated among all parties. In addition, parents were informed in writing about the nature of the study and what their child's participation in it would involve. Finally, in accordance with school district regulations, written permission was obtained from the parents of each of the 48 students who were selected to participate in the study.

During data collection, subjects read two narrative passages on one day and two additional passages on a second day. For each subject, comprehension of each passage was assessed with a different one of the four procedures described in the previous section. The order of administration of test passages and assessment conditions was

counterbalanced to avoid serial effects, to insure that subjects read each test passage under a different assessment condition, and to avoid the confounding of passage and assessment condition. To implement systematically a counterbalanced design of passage and comprehension assessment order, the following steps were taken: (a) definition of four comprehension assessment orders (Order No. 1 = Q, R, C, M; Order No. 2 = R, C, M, Q; Order No. 3 = C, M, Q, R; Order No. 4 = M, Q, R, C), (b) random assignment of four groups to each comprehension assessment order (Group 1 = Assessment Order No. 1; Group 2 = Assessment Order No. 3; Group 3 = Assessment Order No. 2; Group 4 = Assessment Order No. 4; (c) random assignment of 12 subjects to each comprehension assessment order group, (d) definition of four passage orders (Order No. 1 = Passages A, B, C, D; Order No. 2 = Passages B, C, D, A; Order No. 3 = Passages C, D, A, B; Order No. 4 = Passages D, A, B, C), and (e) random assignment of the four passage orders within each assessment order group.

The actual reading of passages and completion of comprehension assessments occurred away from the classroom environment in a distraction-free area of each school building. Data collection generally required from fifteen to twenty minutes per subject on the two consecutive days. Every effort was made to schedule data collection sessions at the same time period each day. No data were collected on

Fridays or on days prior to school holidays. The data collection took place over a five month period, from January through May, 1982.

The testing session. At the beginning of each data collection session, the researcher first established rapport with the subject, explained the study's purpose, and then described the particular comprehension tasks to be done in that session.

To initiate data collection the subject was directed to read his first passage and to engage in the assessment task for that passage according to a prepared script of comments and directions to subjects (see Appendix B). For the questioning assessment, subjects read a given passage silently and responded to the researcher's oral questions at the conclusion of reading. Responses to questions were recorded by the examiner on a prepared record sheet. For the retelling assessment, subjects were recorded retelling what they could remember about a passage. At a later time, the taped retellings were transcribed into typewritten form to facilitate scoring. Specific question probes were not used at the conclusion of subjects' retellings with the exception of "Can you remember anything else?" followed by a waiting period, and then asking the same question a second time. For the cloze assessment, subjects read silently and told the researcher their choice for the missing words in the passage out loud. The researcher recorded subjects'

responses in writing on a second copy of the cloze passage. In addition, field testing showed that a trial cloze passage was needed to alleviate confusion regarding the nature of the cloze task and to relieve anxiety caused by the general unfamiliarity of the process. A trial passage of 129 words was used prior to each cloze assessment.

For the miscue assessment, subjects were recorded reading a passage out loud. To facilitate the quantification and analysis of miscues, the oral reading was played back at a later time for the purpose of marking miscues on a "worksheet" copy of the passage. All miscues were then transferred to a score sheet which also recorded the actual text word for each miscue produced.

At the conclusion of the first day's session, the researcher reviewed with the student the date and time for the second data collection session.

All record sheets necessary to the collection of data are presented in Appendix C. These include the questioning record sheet, the passage content outlines for retelling, the cloze form of passages, and the miscue score sheet.

Data collection procedures were streamlined by the advance preparation of a folder of test materials for each subject. Each folder contained a complete set of passages with their accompanying comprehension assessments and a clear designation of the order in which they were to be administered. Finally, the researcher maintained a record

of reading time for each subject, although this was not an independent variable in the study. By the end of the data collection period, each subject had individually read four passages with the comprehension of each checked under a different assessment condition. In addition, each subject had taken a group-administered standardized reading achievement test which was administered to subjects in small groups near the end of the data collection phase in May.

Hypotheses

Alternative comprehension assessments were administered in this study to determine the nature of relationships among and between alternative measures of reading comprehension. The nature of relationships among measures was investigated by analyzing the data to determine whether specific patterns of association in the results of multiple indicators of comprehension could be identified.

Hypothesis one. The observed data will reflect one of three possible patterns of relationship between and among four indicators of comprehension thereby supporting a specific model of comprehension assessment:

(1) Assessment model I: High correlations will be observed among the results of four measures of comprehension (questioning, retelling, cloze, and miscue). This relationship was tested against the following idealized pattern:

$r_{12} \geq .70$; $r_{13} \geq .70$; $r_{14} \geq .70$; $r_{23} \geq .70$; $r_{24} \geq .70$; $r_{34} \geq .70$.

(2) Assessment model II: Low correlations will be observed among the results of four measures of comprehension (questioning, retelling, cloze, and miscue). This relationship was tested against the following idealized pattern:

$r_{12} \leq .30$; $r_{13} \leq .30$; $r_{14} \leq .30$; $r_{23} \leq .30$; $r_{24} \leq .30$; $r_{34} \leq .30$.

(3) Assessment model III: The correlation observed within comprehension assessment type (an average of the correlations between two product measures and between two process measures) will be higher than is observed across product and process assessment types (an average of the cross-type correlations among the results of four measures of comprehension). This relationship was tested against the following idealized pattern:

$$\frac{r_{12} + r_{34}}{2} > \frac{r_{13} + r_{14} + r_{23} + r_{24}}{4}$$

Hypothesis two. There is no significant linear relationship in a regression model between the dependent variable, results of a standardized comprehension test, and a set of independent variables represented by the results of the four informal measures of comprehension.

Analysis of the Data

In this study, raw scores were the basis for computing measures of relationship between and among alternative indicators of comprehension except in the case of the retelling and miscue assessments where converted raw scores were utilized (percentages of total idea units retold and

adjustment of the total miscue score to provide for numerical comparability among subjects).

To initiate the major analysis, two-way contingency tables were generated and inspected to determine whether non-random factors had operated in the assignment of subjects to assessment task order and passage pattern groups. Chi squares were examined for significance to determine whether extraneous variables requiring statistical control by analysis of covariance could be identified. None were found.

To investigate the nature of relationships among alternative measures of reading comprehension, correlational data were computed. Pairs of scores for each subject included two product measures of comprehension (questioning, retelling), two process measures of comprehension (cloze, miscue), and one standardized measure of comprehension (Gates MacGinitie Reading Test, Level D, Form 3, Houghton Mifflin, 1978). Pearson Product-Moment coefficients of correlation both between and among the four informal measures of comprehension and a standardized measure of reading comprehension, the Gates MacGinitie Reading Test, Level D, Form 3 (Houghton Mifflin, 1978) comprehension subtest, were generated to investigate hypothesis one. A correlation matrix was developed to organize and display the intercorrelations. The calculated coefficients of correlation between and/or among comprehension measures were

considered to be significantly different from zero if the calculated r would occur by chance no more than five in 100 times ($p<.05$).

Finally, multiple correlations were utilized in a regression equation to investigate hypothesis two.

Supplementary Analysis

Introduction. An effort was made to determine whether subjects with the most varied scores on the four measures of comprehension might display distinctive patterns of comprehension performance across the two product and the two process assessment conditions. To identify subjects with the most varied scores, three steps were taken.

First, comprehension assessment data were transformed into a standardized T-distribution (mean of 50; standard deviation of 10) to permit comparison of subjects' relative positions across all measures of reading comprehension examined in the study. Second, subjects whose scores on four comprehension assessments (questioning, retelling, cloze, and miscue) showed the most variability were identified. Third, scores of the most varied subjects were examined to answer the following questions with respect to that group of scores:

Does a particular pattern of scores emerge, particularly with regard to subjects' performance on the product measures (questioning, retelling) as compared to their performance on the process measures (cloze, miscue)? How does a

pattern(s) of scores for the most varied subjects compare to the performance of the whole sample? To what extent does a pattern(s) of scores for the most varied subjects reflect the three proposed models of comprehension assessment under investigation in the study? Do subjects whose scores on the four assessments are the most varied appear similar to and/or different from each other with respect to school, IQ, grade placement, and reading grade level?

The section that follows outlines the specific procedures which were used to identify subjects with the most varied scores. Finally a description of the structure for reporting the researcher's subjective observations of students' behavior and responses during testing concludes this chapter.

Procedure for identifying subjects with the most varied scores. The process for identifying subjects with the most varied scores was initiated by assigning all T-scores represented in the distributions of the questioning, retelling, cloze, and miscue assessments to Quartile One, Two, Three and Four according to the cumulative percent of cases in the frequency distribution for each assessment.

Next, a subject's variability across all four comprehension assessment conditions was determined by examining the quartile placement of each of his four T-scores. Differences in quartile placement among all possible combinations of the four assessments were calculated. Then, the

mean difference in quartile placement among all four assessments was calculated. Figure 1 illustrates the quartile placement for the scores on all four comprehension assessments for Subject A and shows how the differences in quartile placement among all combinations of assessments were obtained. In addition, an example of the calculation of Subject A's mean difference in quartile placement across the four assessments is given.

The final step in identifying subjects with the most varied scores involved generating a new frequency distribution of all mean differences in quartile placement for the entire sample. The highest mean difference figure which most closely corresponded to ten percent of the cases in the new distribution (mean difference of 1.76 quartiles) was considered the cut-off point for purposes of identifying subjects with the most varied scores. Any subject whose calculated mean difference in quartile placement exceeded that cut-off point was considered to have a highly variable score profile.

In other words, subjects with the most variable profiles were those who had an absolute mean difference among their scores on four comprehension assessments which exceeded the cut-off mean difference level of 1.76 quartiles. In this study, five subjects met that criterion. The performance of those five subjects was then analyzed with regard to the questions asked in the introduction to

SUBJECT A

QUARTILE PLACEMENT: (for Subject A's scores on four comprehension measures)

ASSESSMENTS			
Questioning	Retelling	Cloze	Miscue
1	**4**	**3**	**3**

DIFFERENCES IN QUARTILE PLACEMENT AMONG ALL COMBINATIONS OF ASSESSMENTS

Assessments	Difference in Quartile Placement
Between questioning and retelling =	3 Quartiles
Between questioning and cloze =	2 Quartiles
Between questioning and miscue =	2 Quartiles
Between retelling and cloze =	1 Quartile
Between retelling and miscue =	1 Quartile
Between cloze and miscue =	0 Quartiles
	Total 9

MEAN DIFFERENCE IN QUARTILE PLACEMENT: (for Subject A's score on four assessments)

$$\frac{9}{6} = 1.5 \text{ QUARTILES}$$

FIGURE 1. Quartile placement, differences in quartile placement among all combinations of assessments, and calculation of the mean difference in quartile placement based on scores on questioning, retelling, cloze and miscue assessments for Subject A.

this section. A discussion of this analysis is found in Chapter Four.

Subjective observations. The researcher's observations of subject behavior during data collection provided the basis for further supplementary analysis. Behaviors observed included subjects' physical mannerisms during data collection, subjects' overall approach to the assessment tasks, and subjects' responses to assessment stimuli which, in the researcher's judgment, were particularly interesting and/or seemed to reflect their efforts to gain meaning from text. These observations will be presented in Chapter Four.

Summary of Chapter Three

This chapter dealt with the procedures required to undertake a study of the results of alternative assessments of reading comprehension in a sample of learning disabled males. In addition to descriptions of the sample, the stimulus materials, and the four alternative comprehension assessments, this chapter listed the research hypotheses and outlined data collection procedures and the data analysis plan. A procedure for evaluating consistency in each subjects' profile of scores across the four informal indicators of reading comprehension was also outlined. Finally, a structure for organizing the researchers' subjective observations of subject behavior during data collection was described.

CHAPTER 4

Results

Introduction

This study was designed to examine the nature of relationships between and among the results of multiple measures of reading comprehension in a sample of learning disabled students.

First, this chapter will review measures of central tendency and dispersion for the multiple measures of comprehension investigated. Next, analysis of data as it relates to the major hypotheses of the study will be presented. Then, the results of a supplementary analysis of subject variability across the multiple measures of comprehension are discussed. Finally, a summary of the researcher's subjective observations of subjects' behavior during testing and their approach to the assessment tasks is given.

Elimination of Extraneous Variables

No differences in subject performance under randomized specifications for all combinations of passage-and-assessment match were found. This finding was based on inspection of two-way contingency tables to determine whether the random assignment of reading passage pattern and comprehension assessment order were independent of the effects of extraneous variables. Since subject performance did not differ according to order of passages read or according to order of

comprehension assessments accompanying these passages, data were pooled for the major analysis.

Group Means and Standard Deviations for Dependent Measures

Descriptive statistics for the four informal indicators of comprehension and a standardized indicator of comprehension were computed. These statistics were based on a total of five comprehension assessment scores obtained for each subject in the study.

Scores used for all statistical analyses represented (a) the number of questions answered correctly out of a total of 10 questions, (b) the percentage of total idea units retold out of a total ranging from 30 to 38 idea units, depending on the passage, (c) the number of exact replacements out of a total of 25 cloze blanks, (d) the score derived from the degree of semantic acceptability of a total of 25 miscues, and (e) the number of items answered correctly on the comprehension subtest of a standardized reading achievement test. Means and standard deviations for all measures of comprehension investigated are summarized in Table 3.

Normal variability was observed in the distribution of scores for all comprehension assessment conditions. This was apparent in an examination of sample means for the questioning, retelling, cloze, and miscue measures, which showed that the range between one standard deviation above

TABLE 3

MEAN, STANDARD DEVIATION, SKEWNESS, KURTOSIS, MINIMUM, AND MAXIMUM FOR FOUR COMPREHENSION ASSESSMENTS

Comprehension Assessment	Mean	Standard Deviation	Skewness	Kurtosis	Min.	Max.
Questioning*	7.750	1.695	-0.547	-0.309	4.000	10.000
Retelling *	23.896	8.161	0.166	-0.262	6.000	41.000
Cloze *	12.271	3.636	-0.275	-0.072	2.000	18.000
Miscue *	55.854	14.951	-0.742	-0.674	24.000	74.000
Standardized Test	24.063	9.325	0.184	-0.735	4.000	41.000

Note: Study Sample = 48.

* Maximum possible scores for each informal assessment were Questioning (10 points), Retelling (100 percent of total idea units retold), Cloze (25 points), Miscue (75 points).

and below the mean in each distribution contained approximately two-thirds of all scores in the sample. Normal variability was also observed in the approximate percentage of subjects in the sample who had scores within one of four quartile categories represented in the distribution of scores for each assessment condition. Figure 2 shows that for the retelling, cloze, and miscue assessments, the percentage of subjects with scores in each of the four quartile categories was relatively even.

This pattern of an even distribution of scores among four quartiles was only slightly different for the questioning assessment. The Supplementary Analysis section of Chapter Three explains the procedure used for determining the quartile placement of scores in the distributions of the questioning, retelling, cloze, and miscue assessments.

Subject performance on four informal measures of comprehension. Among the product indicators of comprehension (questioning and retelling), subjects answered nearly 80 percent of the 10 questions asked, regardless of which narrative passage was read. Eight subjects made the highest score for this assessment (10 points for 10 questions answered correctly), while another 10 subjects received nine points for answering nine questions correctly. The questioning assessment task involved answering not only textually explicit questions but also answering textually implicit and scriptally implicit questions.

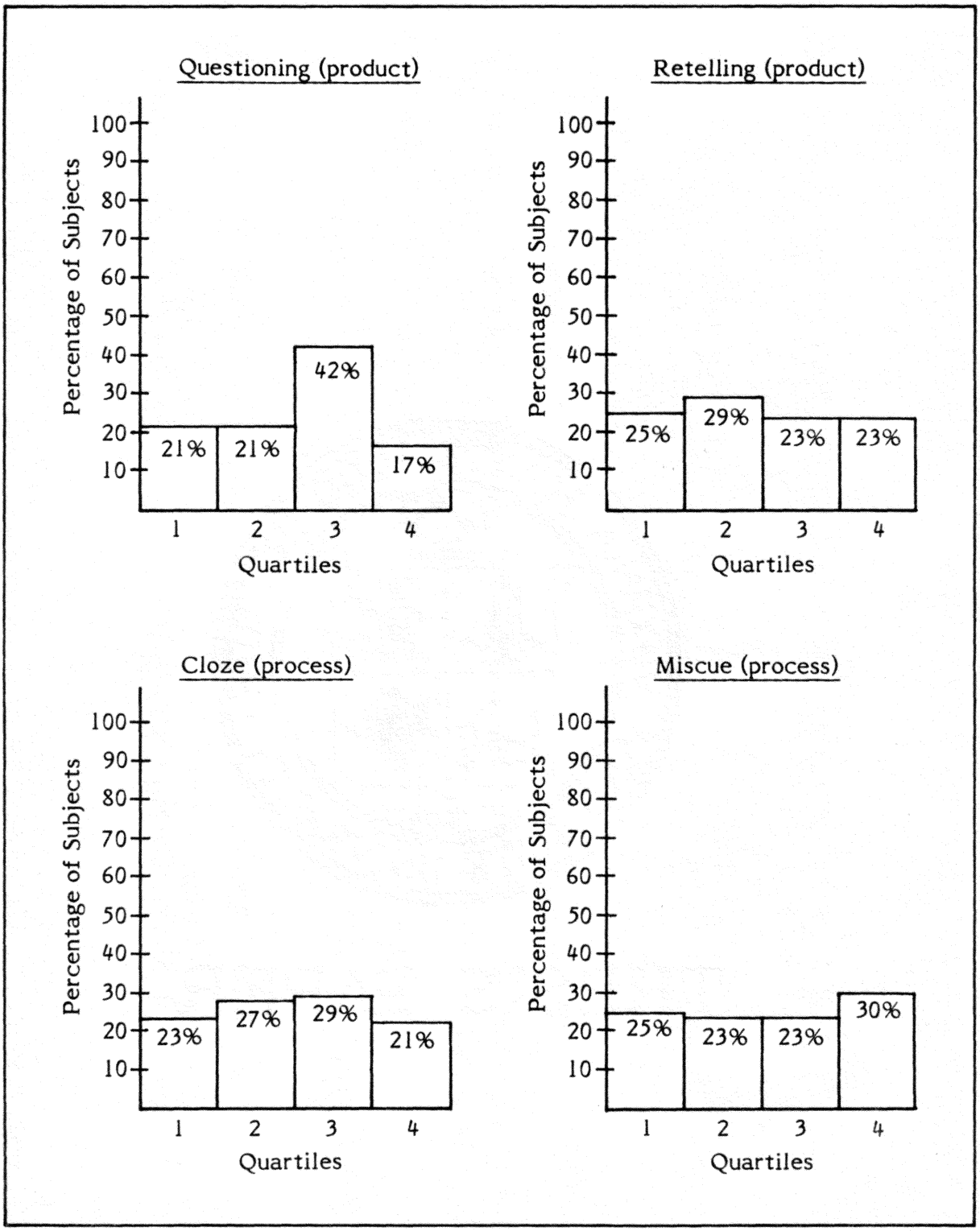


FIGURE 2. Percentage of subjects with scores in each of four quartiles (Q_1=0%-25%; Q_2=26%-50%; Q_3=51%-75%; Q_4=76%-100%) by comprehension assessment.

For the retelling assessment, subjects retold a mean of 24 percent of the total number of idea units contained in all passages. The maximum percentage of idea units retold was 41 percent, with only two subjects scoring at that level. When main ideas and details were separated, subjects retold an average of 32 percent of the main ideas and 22 percent of the details. Table 4 contrasts the total number of possible idea units for each passage with the mean retelling score for that passage. For this comparison, main ideas and details were not separated.

An examination of Table 4 shows that Passage C had the lowest mean percent of idea units recalled of any of the other passages (17.75). However, while Passage A contained nearly the same number of idea units as Passage C, subjects who read Passage A retold approximately 26 percent of its idea units in comparison to the 17.75 percent of idea units retold by subjects who read Passage C. The mean percent of idea units retold for Passages B and D were also approximately 26 percent. Except for the slight variation noted in the retellings of Passage C, subjects tended to retell about one quarter of the total idea units regardless of which passage was read. This finding was consistent with findings reported in the literature. The highest score for this assessment was 41 percent of the total idea units retold.

Among the process indicators of comprehension (cloze and miscue), a mean score of 12 points on the cloze

TABLE 4

A COMPARISON OF MEAN RETELLING SCORES ACROSS TEST PASSAGES

Test Passage	Cases	Maximum Total Idea Units	Mean Percent of Total Units Retold
Passage A - "Give the Ball to Wilt"	12	30	25.9
Passage B - "Harriet Tubman"	12	38	26.3
Passage C - "First in the Sky"	12	31	17.75
Passage D - "Fine Animal Gorilla"	12	34	25.5

assessment indicated that subjects produced exact replacements for nearly one-half (48 percent) of the total of 25 cloze blanks across all passages, based on an every-tenth-word deletion pattern. The highest score achieved on this assessment was 18 points.

Twenty-five oral reading miscues of each subject were analyzed for semantic acceptability, with adjustments for self-corrections in the scoring process. The highest score for the assessment was 74 out of 75 points, achieved by two subjects. The mean score for the assessment was 56 points.

Subject performance across multiple measures of comprehension. Figure 3 provides a summary comparison of subjects' comprehension performance across the questioning, retelling, cloze, and miscue assessments. Because each assessment had a different maximum score, the comparison was based on a conversion of the mean for each assessment to a percentage of that maximum score. On this basis, overall performance on the retelling assessment was the lowest of the comprehension measures investigated in this study.

Time. The time that subjects needed to read the narrative passages during data collection was noted, even though it was not a variable of importance in the study. For the passages that were followed by either a questioning, retelling, or miscue comprehension assessment, subjects took approximately three-and-one half minutes to finish the necessary reading. For the cloze assessment, however,

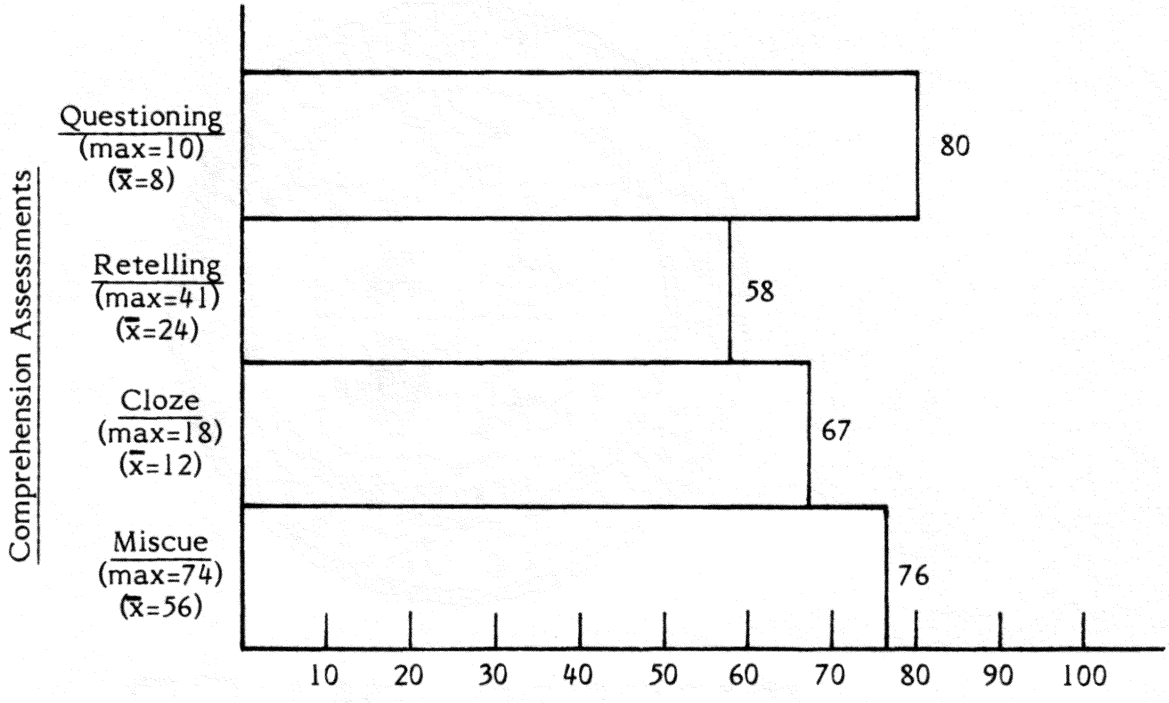


FIGURE 3. Comparison of the sample's relative comprehension performance based on a conversion of the mean for each assessment to a percentage of the maximum score possible.

subjects needed an average of 10 minutes to complete the accompanying passage because the cloze assessment not only required silent reading but also the oral completion of cloze blanks. Sixteen subjects took longer than 10 minutes to read the cloze passage, with one subject requiring a total of 20 minutes to finish.

Hypothesis One

Introduction. Hypothesis one involved determining whether the data collected reflected one of three proposed patterns of relationship among four indicators of comprehension thereby supporting a specific model of comprehension assessment. The patterns of relationship tested were:

(1) High correlations ($\geq .70$) observed among the results of four measures of comprehension (questioning, retelling, cloze, and miscue), thereby supporting assessment model one.

(2) Low correlations ($\leq .30$) observed among the results of four measures of comprehension (questioning, retelling, cloze, and miscue), thereby supporting assessment model two.

(3) Higher correlations ($\geq .70$) observed in the results of comprehension measures within assessment type (product measures versus process

measures) than across assessment type ($\leq .30$), thereby supporting assessment model three.

Methodology. To test the hypothesis, Pearson Product-Moment coefficients of correlation (zero-order) were computed for pairs of variables expressed as comprehension assessment raw scores. The intercorrelations among variables were organized in the standard matrix format displayed in Table 5. In addition, an average of the correlations within two assessment types, product and process, was computed and then compared to an average of across-type correlations.

Analysis. The hypothesized pattern of higher correlations among the results of four measures of comprehension (questioning, retelling, cloze, and miscue) was tested by the following formula:

$r_{12} \geq .70$; $r_{13} \geq .70$; $r_{14} \geq .70$; $r_{23} \geq .70$; $r_{24} \geq .70$; $r_{34} \geq .70$.

Table 5 reveals that the highest relationship among the four comprehension measures was that observed between the results of the questioning and cloze assessments, r=.48 (p=.000). All other correlations observed among measures were lower in magnitude, ranging from a zero correlation between the results of the retelling and miscue assessments of r=-.0252 (p=.432) to a relationship of r=.46 (p=.000) between the results of the cloze and miscue assessments.

Therefore, these data are not consistent with the model that postulates a high correlation between the four assessments used in this study.

TABLE 5

ZERO-ORDER CORRELATIONS OF ALTERNATIVE COMPREHENSION ASSESSMENTS (N=48)

	SLQS[a]	SRSTOTS[b]	SCS[c]	CTOTORAL[d]	GMRC[e]
SLQS	1.0000 (48) P=*****	.2996* (48) P=.019	.4773* (48) P=.000	.3285* (48) P=.011	.3147* (48) P=.015
SRSTOTS		1.0000 (48) P=*****	-.0012 (48) P=.497	-.0252 (48) P=.432	.2542* (48) P=.041
SCS			1.0000 (48) P=*****	.4625* (48) P=.000	.5366* (48) P=.000
CTOTORAL				1.0000 (48) P=*****	.5815* (48) P=.000
GMRC					1.0000 (48) P=*****

[a]Questioning Assessment

[b]Retelling Assessment

[c]Cloze Assessment

[d]Miscue Assessment

[e]Standardized Test

*p <.05

A second hypothesized pattern, low correlations among the results of four measures of comprehension (questioning, retelling, cloze, and miscue), was tested with the following formula:

$r_{12} \leq .30$; $r_{13} \leq .30$; $r_{14} \leq .30$; $r_{23} \leq 30$; $r_{24} \leq .30$; $r_{34} \leq .30$.

Table 5 reveals that the hypothesized pattern of low correlations among the results of four measures of comprehension was given mixed support. The lowest relationships among the four measures were apparent in the retelling and cloze scores (r=-.0012; p=.497) and in the retelling and miscue scores (r=-.0252; p=.432). These relationships were not statistically significant at p= <.05. The lowest positive correlation was between the questioning and retelling scores (r=.30; p=.019). All other correlations observed among measures were above r=.30 in magnitude, ranging from a low of r=.33 (p=.011) for questioning and miscue assessments to a high of r=.48 (p=.000) for questioning and cloze assessments.

Therefore, while these data are not entirely consistent with assessment model two, in which there would be no r value greater than .30, two of the assessments (retelling-cloze, retelling-miscue) have virtually a zero correlation, consistent with an absence of relationship. In addition, two other relationships (questioning-retelling; questioning-miscue) ranged at or near .30 (consistent with model II), and a near-moderate correlation for two other pairs of measures (questioning-cloze; cloze-miscue) was noted.

This pattern of low and near-moderate positive correlations tend toward assessment model two for two reasons. First, the coefficients of determination (Hinkle, Weirsma, & Jurs, 1979, pp. 85-87) of these relationships show a high percentage of unaccounted for variance among the questioning-retelling, questioning-miscue, questioning-cloze, and cloze-miscue measures. Second, a simple average of all possible coefficients of correlation among the four measures is .26, well within the hypothesized r of $\leq .30$.

In short, as between assessment model one and assessment model two, these data are more nearly consistent with the latter than with the former. Relationships among measures suggest, however, that results of the questioning, cloze, and miscue assessments appear to cluster together while comprehension as measured by the retelling assessment appears to be noticeably different. Further, the overall pattern of low relationships among measures suggest that their results cannot be considered completely interchangeable.

Finally, the hypothesized pattern of higher correlations ($r = \geq .70$) observed within assessment type (product measures versus process measures) than across assessment types (among the results of four informal assessments, $r \leq .30$) was tested with the following formula:

$$\frac{r_{12} + r_{34}}{2} > \frac{r_{13} + r_{14} + r_{23} + r_{24}}{4}$$

Table 5 reveals that the hypothesized pattern was observed, in part. First, the association between the two product measures of comprehension, questioning and retelling, was r=.30 (p=.019). Within the process assessment category, the association observed between the cloze and miscue measures of comprehension was r =.46 (p=.002). Indeed, as Table 6 demonstrates, the coefficients of determination show that the relationship between the cloze and miscue process measures accounts for more than twice as much of the variance between them as the relationship between the questioning and retelling product measures does with respect to variance accounted for between them (21 percent versus 10 percent). This factor may suggest more similarity between process measures than between product measures in this study, but the overlap between variables was not enough to suggest that the comprehension behaviors required by product assessments are completely different from those required by process assessments, however. Next, the average of the two within-type correlations (.38) was higher than the average of the across-type correlations (.20), consistent with the hypothesis of assessment model III that this pattern would be observed in the data.

Despite the higher association between process than between product measures as well as within assessment type than across assessment types, the observation of model III in the data is problematic for two reasons. First, the

TABLE 6

SUMMARY OF RELATIONSHIPS OBSERVED WITHIN COMPREHENSION ASSESSMENT CATEGORIES

Assessment Category	r	p	Coefficient of Determination	Percent of Variance Explained
Within Product Measures:				
Questioning-Retelling	.2996	0.019	0.0987	Approx. 10%
Within Process Measures:				
Cloze-Miscue	.4625	0.000	0.2139	Approx. 21%

averaging of across-type relationships masks the fact that the questioning assessment actually has a higher correlation with either cloze or miscue than it has with the other product measure, retelling (r=.48 and r=.33 with cloze and with miscue versus r=.30 with retelling). In fact, the highest correlation recorded in the study is between the product measure, questioning, and the process measure, cloze (r=.48), with the greatest amount of variance being accounted for in the study (approximately 23 percent) in the distributions of the questioning and cloze assessments (see Table 7). Second, the across-type average was skewed by the zero correlations (r=-.0012 and r=-.0252) of retelling with the cloze and miscue measures. Therefore, even the slightest support that the higher within-type average of correlations gives to assessment model three is likely to be spurious in comparison to the across-type average.

In short, assessment model three, which would logically imply that comprehension as product and comprehension as process require different assessments, is not strongly supported by the data. Despite the higher association of measures observed in the average of correlations within assessment type (product-process) than in an average of four informal measures across both types, the correlations between product and between process measures were not so high as to justify the assertion that abilities measured by questioning and retelling are altogether different from abilities measured by cloze and miscue.

TABLE 7

SUMMARY OF RELATIONSHIPS OBSERVED ACROSS COMPREHENSION ASSESSMENT CATEGORIES

	Process Measures							
	Cloze				Miscue			
Product Measures	r	p	Coeff. Deter.	% Var. Expl.	r	p	Coeff. Deter.	% Var. Expl.
Questioning	.4773	0.000	0.2278	Approx. 23%	.3285	0.011	0.1079	Approx. 11%
Retelling	-0.0012	0.497	0.000	None	-0.0252	0.432	0.000	None

Summary concerning hypothesis one. Data were collected to determine whether one of three proposed patterns of relationship among four indicators of comprehension would be observed, thereby supporting a specific model of comprehension assessment.

Based on an examination of intercorrelations between the results of questioning, retelling, cloze, and miscue comprehension assessments, the analysis of data reveals the following:

First, these data were not consistent with the high associations among measures postulated in model one. There was no correlation in this study greater than or equal to .70; indeed, there was no correlation greater than or equal to .50 in this study.

Second, not withstanding the above, the data showed mixed support for model two, that there would be no correlations greater than .30, in the sense that the average of all possible correlations was less than .30, and in the general sense that half the correlations examined were below .30 and the rest were below .50. There were, however, near moderate correlations for some variables.

Third, the data supported model three in the limited sense that an average of across-type correlations was lower (.20) than an average of within-type correlations (.38). Certain correlations did not fit the proposed pattern, however, with the highest correlation between informal measures

in the study crossing assessment types (questioning and cloze r=.48, p=.000), a finding directly opposed to the assumption of assessment model three that product and process measures would be dissimilar.

In summary, the data do not confirm any of the three models of comprehension assessment but are more consistent with a model that viewed retelling as requiring one cluster of comprehension behaviors and questioning, cloze, and miscue measures as requiring another cluster. Even so, the relationships among the informal assessments examined in this study are not of a high enough magnitude to confirm their interchangeability in measuring the reading comprehenison of LD subjects.

Relationship of informal comprehension assessments to a standardized reading measure. Table 5 reveals that 11 correlations obtained between informal comprehension assessments and a standardized measure (Gates MacGinitie Reading Test, Level D, Form 3, Houghton Mifflin, 1978) were significant at an alpha level of .05 or less. Overall, the two process comprehension assessments were more highly associated with the standardized measure (cloze r=.54, p=.000; miscue r=.58, p=.000) than were the two product comprehension assessments (questioning r=.31, p=.015; retelling r=.25, p=.041).

These results indicate that the four comprehension measures examined in this study shared variance with the

variance observed in the standardized test, but to a greater extent for the cloze and miscue measures than for the questioning and retelling measures. The lower relationships observed between the questioning and retelling assessments and the standardized test may have been partly attributable to variables such as differences in the total number of items on the questioning measure as compared to the standardized test and to other differences in the task demands of the informal and the standardized measures described in Chapter Two. None of the correlations observed between these indicators of comprehension were high enough to suggest that one measure could be substituted for the standardized measure in assessing a reader's comprehension, however.

Hypothesis Two

Introduction. Hypothesis two was concerned with which of the four reading comprehension assessments (questioning, retelling, cloze, and miscue) most closely agreed with subjects' performance on a standardized, group administered reading comprehension test. Specifically, the null hypothesis stated that there would be no observed linear relationship in a regression model between the dependent variable, results of a standardized comprehension measure (Gates MacGinitie Reading Test, Level D, Form 3, Houghton Mifflin, 1978) and a set of independent variables, the results of two product measures (questioning, retelling) and the results of

two process measures (cloze, miscue) of comprehension. If a linear relationship were observed in the data, then an identification of the assessments which best predicted standardized test performance could be made.

Methodology. To test the null hypothesis, multiple correlations were utilized in a step-wise regression analysis to determine which of the four independent variables, singly or in combination, best predicted subject performance on the standardized test. The percentage of variance on the criterion measure as explained by the results of alternative comprehension assessments was also investigated by examining the square of the multiple correlation coefficient (Hinkle, Weirsman & Jurs, 1979, pp. 368-389). The cumulative percentage of variance explained was the main focus of the analysis.

Analysis. The null hypothesis of no observed linear relationship between the dependent variable, standardized test, and the four alternative comprehension assessments was rejected at $p = <.05$ for three out of the four assessments. In order of their contribution to the explanation of variance, the independent variables represented by the (a) miscue, (b) cloze, and (c) retelling assessments were selected by the step-wise procedure to account for successively greater amounts of variance in comprehension performance as measured by the standardized test (see Table 8). Questioning (product measure) was the variable excluded from the equation.

TABLE 8

STEP-WISE MULTIPLE REGRESSION OF FOUR ALTERNATIVE COMPREHENSION ASSESSMENTS ON A STANDARDIZED TEST OF READING COMPREHENSION

Step No.	Variable Entered	Mult. R	SE Est.	R^2	R^2 Change	B	SE B	F
1	Miscue	0.58152	7.66811	0.33816	0.33816	0.36269	0.07481	23.504
2	Cloze	0.66520	7.19938	0.42929	0.09112	0.26447	0.07922	11.144
3	Retelling	0.70692	6.81655	0.49974	0.07045	0.26967	0.07504	12.915

Specifically, the relationship between the independent variable miscue (process measure) and the dependent variable (standardized comprehension test) was the highest (r=.58) of any of the independent variables examined. Approximately 34 percent of the variance in the dependent variable was accounted for by factors measured on the miscue indicator of comprehension. At the second step in the equation, the Multiple R rose to .66 following the inclusion of the independent variable cloze (process measure), bringing the cumulative percentage of variance in the standardized test accounted for by factors in miscue and cloze to nearly 43 percent. At the third step, inclusion of the independent variable retelling (process measure) brought the Multiple R to .71, thus accounting for nearly 50 percent of the variance observed in the dependent variable.

Summary concerning hypothesis two. The data are not consistent with the null hypothesis, but the independent variables (miscue, cloze, and retelling assessments) identified in the regression analysis only accounted for approximately 50 percent of the variance in the dependent variable, the standardized test. So, even after the regression analysis, there remains a substantial amount of variance in the standardized test in this study that is not accounted for by the miscue, cloze and retelling assessments. It should be noted that the standardized test differed from the informal comprehension assessments in being a group, paper-and-pencil test.

The first predictor chosen in the regression analysis was miscue because it had the highest positive correlation with the standardized test of any of the other four measures, r=.58; p=.000. This was the highest correlation between any two factors obtained in the study, accounting for 34 percent of the differences in subjects' standardized test and miscue assessment performance.

The variable entered at the second step of the equation was cloze, adding an additional nine percent to the total amount of variance accounted for among measures. While the miscue and cloze assessments are identified in this study as process measures of comprehension, their higher associations with the standardized measure may be apparent, in part, because the cognitive processes involved in all three tasks are not dissimilar.

Retelling was the last independent variable identified by the regression program, by virtue of the fact that it accounted for a larger portion of the remaining variance than did the other remaining predictor variable. The retelling assessment did not show much association with any other assessment investigated in the study. Even so, its entry into the regression equation and its relationship to other assessments suggest that retelling may have some unique role in accounting for differences in student comprehension performance.

Furthermore, the questioning assessment does not appear from the regression analysis to account for any significant portion of the remaining variance beyond that accounted for by miscue, cloze and retelling. Many r-values associated with questioning cluster around r=.30 (retelling: r=.30; cloze: r=.48, the major exception; miscue: r=.33; and standardized test: r=.31). Nevertheless, the questioning assessment may be considered a factor in explaining the results of the standardized measure even though it did not appear in any step of the regression equation. Questioning apparently overlaps with other variables that were identified in the regression analysis and that better account for the variance observed between formal and informal measures of comprehension.

In brief, no independent variable by itself had more than a moderate correlation with the standardized test. It is only when the miscue, cloze and retelling assessments were considered in combination that they suggested a higher association with the standardized test of .71. Yet a large portion of the variance associated with the standardized test remained unaccounted for by variance in the miscue, cloze, and retelling assessments.

Supplementary Analysis

Summary of the most varied cases. In accordance with procedures outlined in Chapter Three, five subjects were identified as having the most varied scores in the sample.

These subjects represented 10.41 percent of the sample. Examining the complexities of the assessment task demands of the four informal measures investigated in this study was believed necessary in subjects with the most varied profiles because their actual comprehending behavior was not clearly apparent in the strictly quantitative aspect of their scores.

The single descriptive characteristic held in common by these subjects was their current level of performance in reading (as measured by a standardized test other than the Gates MacGinitie Reading Test). All of them performed within the third-to-fifth reading grade level range.

In contrast, three of the five subjects were eighth graders and two were seventh graders. In addition, three of them had WISC (Wechsler, 1974) Full Scale IQ scores within the average range of intellectual functioning (90-109), while two students had Full Scale IQ scores in the low average range (80-89).

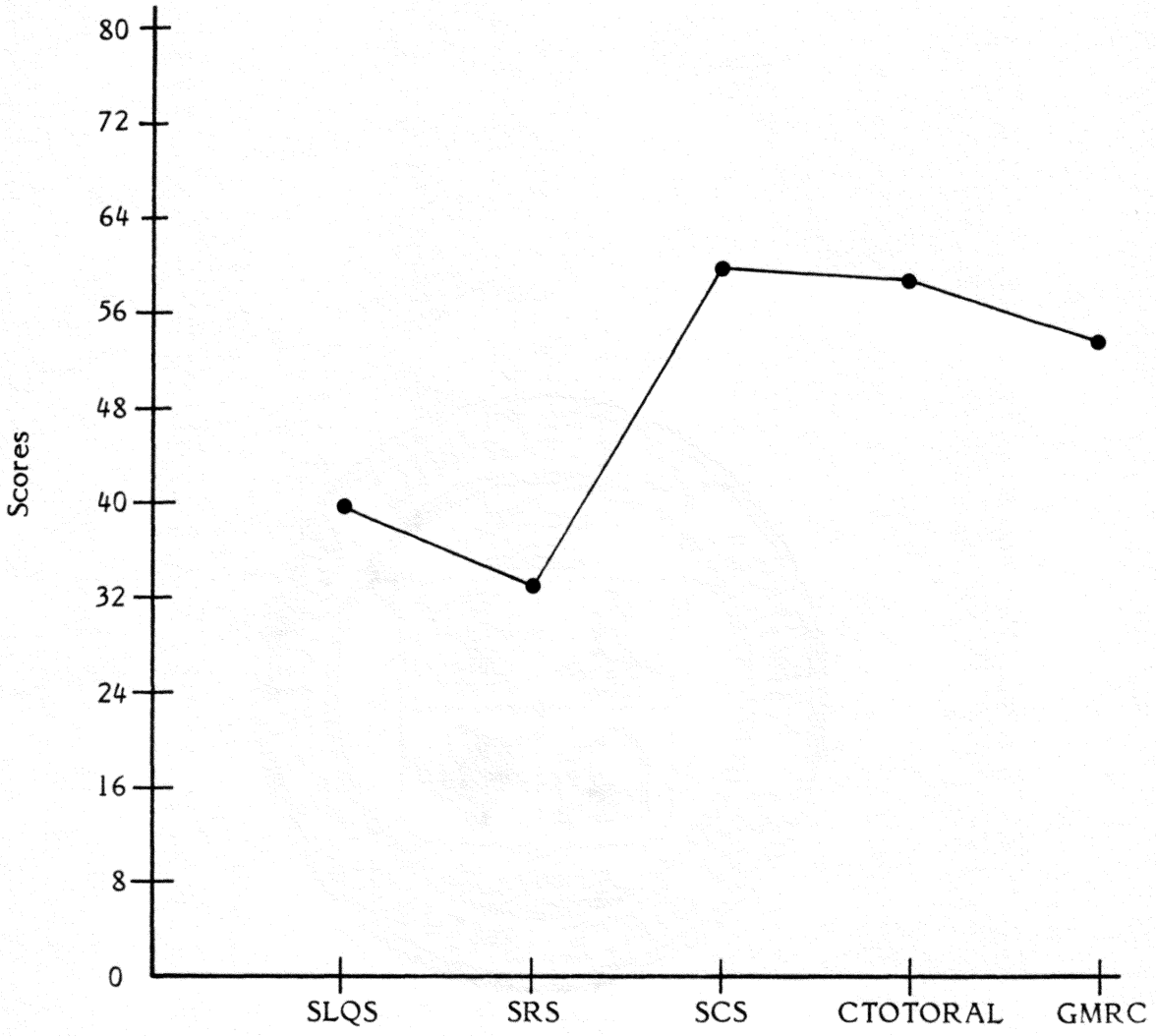
Patterns of performance across assessment conditions. Figures 4 to 8 illustrate the variability in these five subjects' performance across the comprehension assessments investigated in the present study. Two patterns of scores within the five profiles were identifiable. In each case where a subject's scores were among the most varied in the sample, he scored either (a) low on product and high on process measures, or (b) high on product and low on process

measures. The following sections describe subject performance within the two scoring patterns identified.

Low product scores/high process scores. The first pattern of variability in scores is represented by low scores on the two product measures (questioning, retelling) and high scores on the two process measures (cloze, miscue). Three out of the five most varied cases demonstrated this pattern. (See Figures 4, 5, and 6.)

On the questioning assessment, each of the three subjects who demonstrated this pattern had scores that were one standard deviation below the mean for the sample, or a total of six questions correct out of 10. One subject in the group, subject 3, seemed to have particular trouble remembering the questions that were asked of him long enough to construct a response. He also requested numerous repetitions of questions before answering.

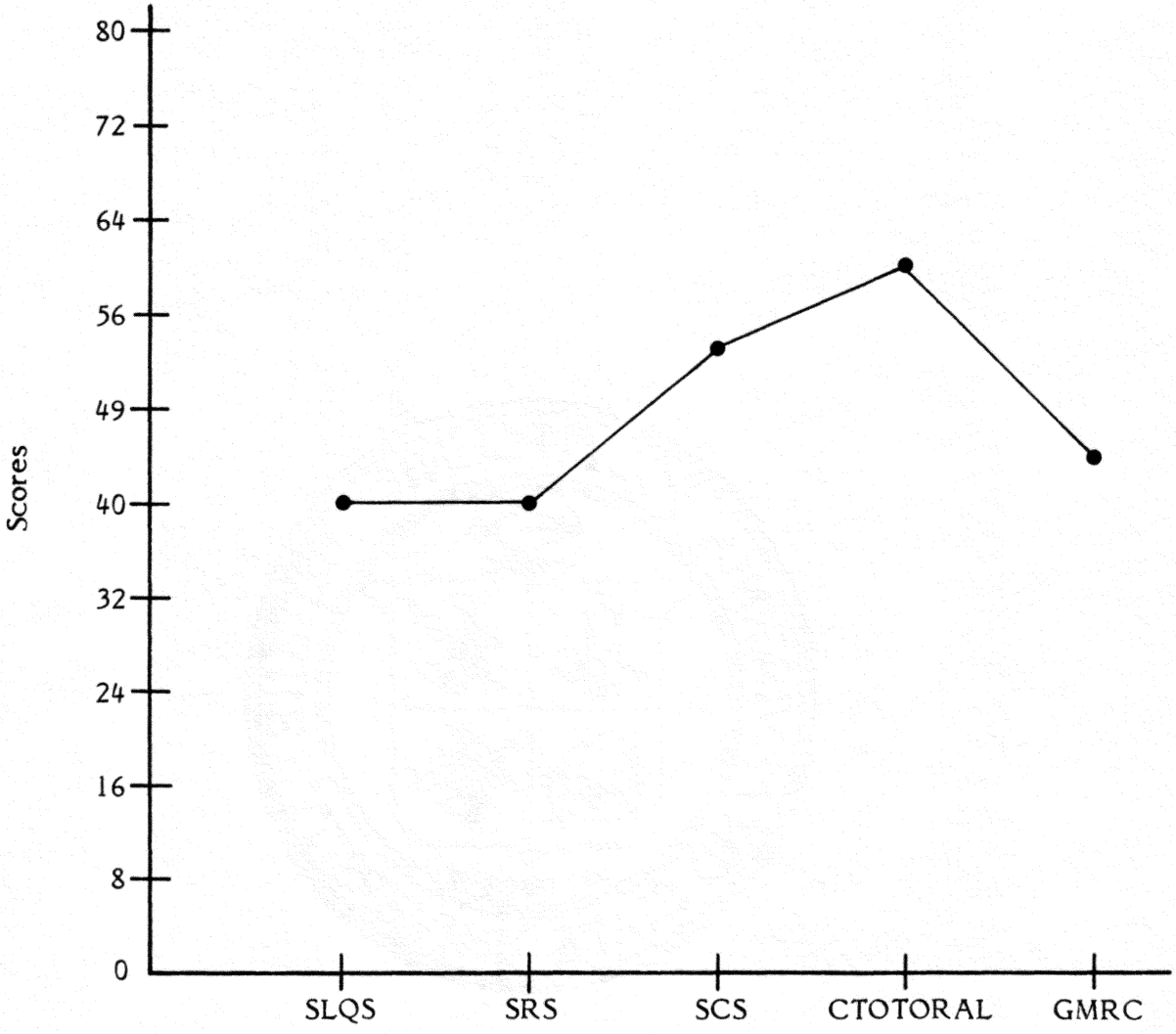
Subject 12's score was low on the questioning assessment because he could not answer either of the two questions which required the application of background knowledge and experience. In addition, his answers to the remaining questions were short and unelaborative. On occasion subject 12 came very close to producing a correct response, but then included one word in his answer which caused the investigator to infer a lack of understanding. For example, in a question asking how one could tell from the story that Wilt Chamberlain's father cared about him, subject 12 replied, "He tried to raise the ceilings because Wilt was too



Alternative Reading Comprehension Assessments

SLQS -	Questions Assessment
SRS -	Retelling Assessment
SCS -	Cloze Assessment
CTOTORAL -	Miscue Assessment
GMRC -	Standardized Assessment

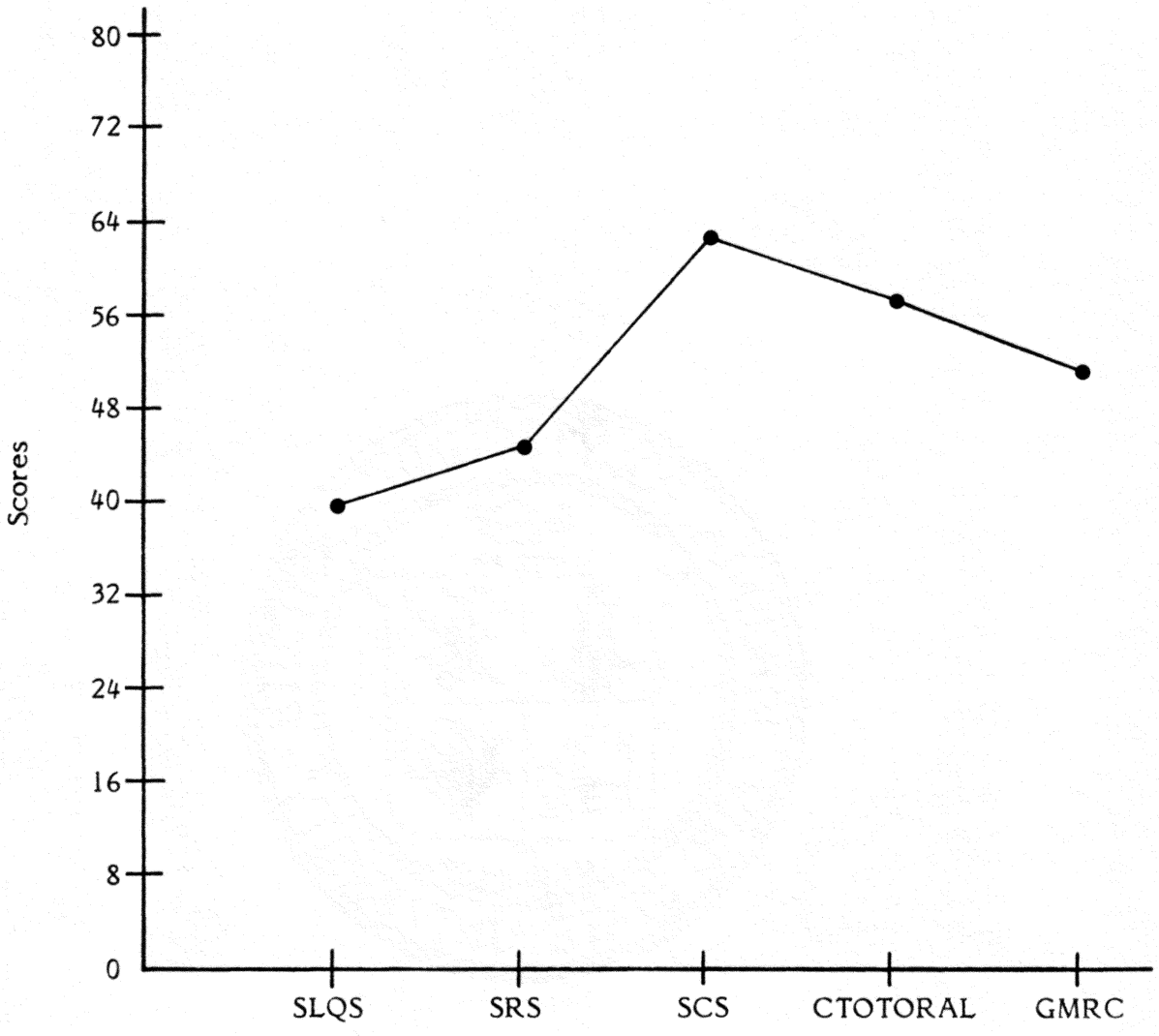
FIGURE 4. Comprehension assessment profile for subject 03, low on product, high on process measures.



Alternative Reading Comprehension Assessments

SLQS -	Questions Assessment
SRS -	Retelling Assessment
SCS -	Cloze Assessment
CTOTORAL -	Miscue Assessment
GMRC -	Standardized Assessment

FIGURE 5. Comprehension assessment profile for subject 12, low on product, high on process measures.



Alternative Reading Comprehension Assessments

SLQS - Questions Assessment
SRS - Retelling Assessment
SCS - Cloze Assessment
CTOTORAL - Miscue Assessment
GMRC - Standardized Assessment

FIGURE 6. Comprehension assessment profile for subject 35, low on product, high on process measures.

tall." (The correct response to the question was that Wilt's father raised the lights that hung from the ceilings so that Wilt would not bump into them). Without further questioning, it is difficult to know whether this subject's experience with "raised ceilings" had caused him to respond in this way or whether he had actually comprehended the text relating to the raised lights but had simply failed to express his comprehension clearly. Another subject, although not one of the five with the most varied scores, made a similar reference to raising ceilings during a retelling of content about the same passage. He said, "His Dad had to . . . higher the ceiling lamps because he didn't know a way how to higher the ceiling."

The last of the three low-product subjects, subject 35, answered the two scriptally implicit questions about his passage correctly but missed all but one of the four textually implicit questions. This type of question-answering behavior might lead an examiner to conclude that while the subject integrated his background knowledge and experience into the understanding of text information, he was nevertheless unable to integrate various pieces of information within the text to generate an answer that was not explicitly stated there. Subject 35's responses to the textually explicit questions were all correct, except one.

On the second product measure, retelling, the same subject who had trouble remembering questions, subject 3,

also had a difficult time getting his retelling underway. In addition, he constantly squirmed and rubbed his hands on his face and head during the assessment process and paused frequently between each idea unit retold. Nevertheless, in his retelling of passage content about a woman airline pilot, an interesting construction of language was observed in his reference to her as a "woman flighter" instead of as a "pilot." The words "flight" and "flying" appeared frequently in the passage and perhaps were in his mind when he was trying to recall the word "pilot" but could not find it.

Similar physical manifestations of anxiety were observed in subject 12 during retelling. This subject even verbalized the difficulty he had with retelling to the researcher.

Subject 35 had a retelling protocol that contained a considerable amount of information, but much of it was not tied together in a logical way. This factor complicated the scoring of the protocol and limited the awarding of credit for some of the idea units retold.

Overall, each of these three subjects with a pattern of low product and high process scores retold a slightly higher percentage of main idea units than detail idea units. Subject 35 retold 20 percent of the total idea units contained in the passage he read, while subjects 12 and 3 retold sixteen percent and ten percent of the total idea units contained in their passages respectively. The mean percentage

of total idea units retold across all passages for the whole sample was 24 percent.

With regard to the cloze and miscue (process) measures of comprehension investigated in this study, the scores of these three subjects were considerably higher than their scores on the product measures. On the cloze assessment subjects 35 and 3 had a score of 17 and 16 respectively, at least one standard deviation above the mean of the sample (which was 12 exact replacements out of a total of 25 cloze blanks). For both of these subjects, the cloze score was their highest among all of the other comprehension assessments administered.

Although cloze scoring guidelines did not count synonym replacements as correct, each of these three subjects with higher process scores made between three-to-five appropriate synonymous replacements of cloze blanks, reflecting their ability to use language in a positive way. Subject 12, whose expressive language was not elaborative on the questioning assessment, produced the lowest cloze score in this group of three subjects, but he also made the highest number of synonym replacements (5). Had his synonymous responses been counted correct, his score for the cloze assessment would have gone up from a total of 13 points, near the sample mean, to a total of 18 points, well above the sample mean.

In addition, Subject 3 completed the cloze assessment in eight minutes with subject 35 and 12 completing it in 14 and 10 minutes, respectively. The mean completion time for the sample for the cloze assessment was 10 minutes.

Finally, for the miscue assessment, all three subjects made a grand total of three miscues each which, according to the scoring guidelines, resulted in total scores for the assessment ranging from 68 to 71 points out of a maximum of 75 points. For the most part, these scores were nearly two standard deviations higher than the mean for the sample on the miscue asssessment of 56 points. Specifically, subject 12's score of 71 points was exactly two standard deviations above the sample mean. All three of his miscues retained meaning, even though one reflected an error in subject-verb agreement. The outstanding feature of subject 12's oral reading was that he read the name of a gorilla named Kolo as "Kol<u>a</u>" each of the 12 times it appeared in the passage. "Kola" seemed to be the word the subject was comfortable using as the gorilla's name and he made no attempt at a correction. The letters "o" and "a" are also graphically similar which may have contributed to the miscue. In addition, "Cola" is a familiar English proper noun, Kolo is not.

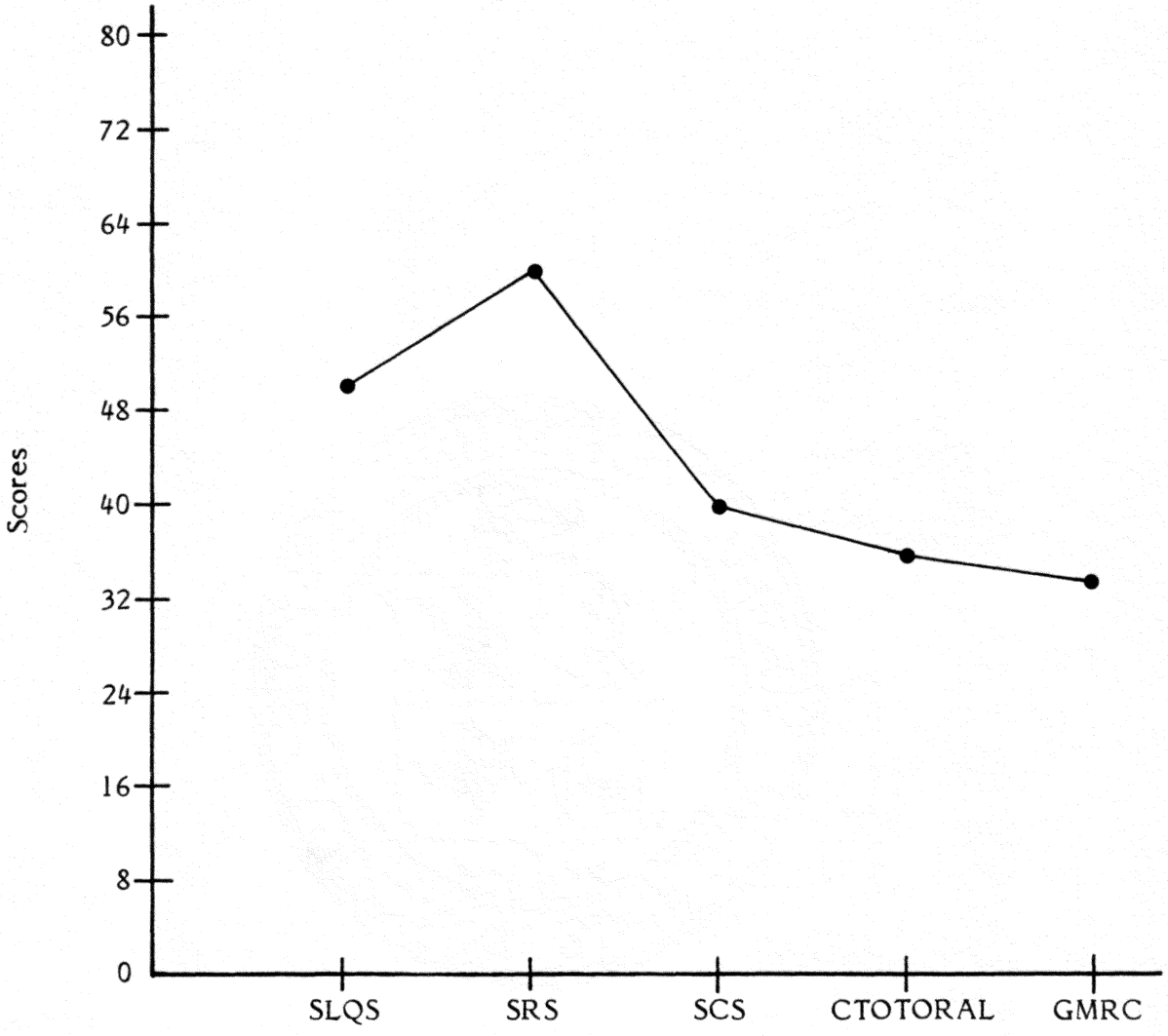
None of subject 35's three miscues retained meaning, however, and two of subject 3's three miscues also did not. An interesting reflection of the impact of a possible lack of experience in vocabulary development was observed in

subject 35's referral to a "727 jetliner" (text language) as a "seven hundred-and-twenty-seven jetliner." In all probability, this reference would not have affected overall understanding of the passage's reference to an airplane since he read the word "jetliner" without difficulty. Nevertheless, it does show his unfamiliarity with the labels for various types of commercial aircraft.

High product scores/low process scores. The second pattern of variability in comprehension assessment scores is represented by high scores on the two product measures (questioning, retelling) accompanied by comparatively lower scores on the two process measures (cloze, miscue), exactly the opposite of the first pattern identified. Two out of the five most varied cases demonstrated this pattern. (See Figures 7 and 8.)

For both of these subjects, the retelling score was easily the highest score among the four comprehension assessments. Subject 40's score of 32 percent of the total passage idea units retold was exactly one standard deviation above the mean for the sample (24 percent total idea units retold) while subject 52's retelling of 40 percent of the total idea units in his passage was just over two standard deviations above the sample mean.

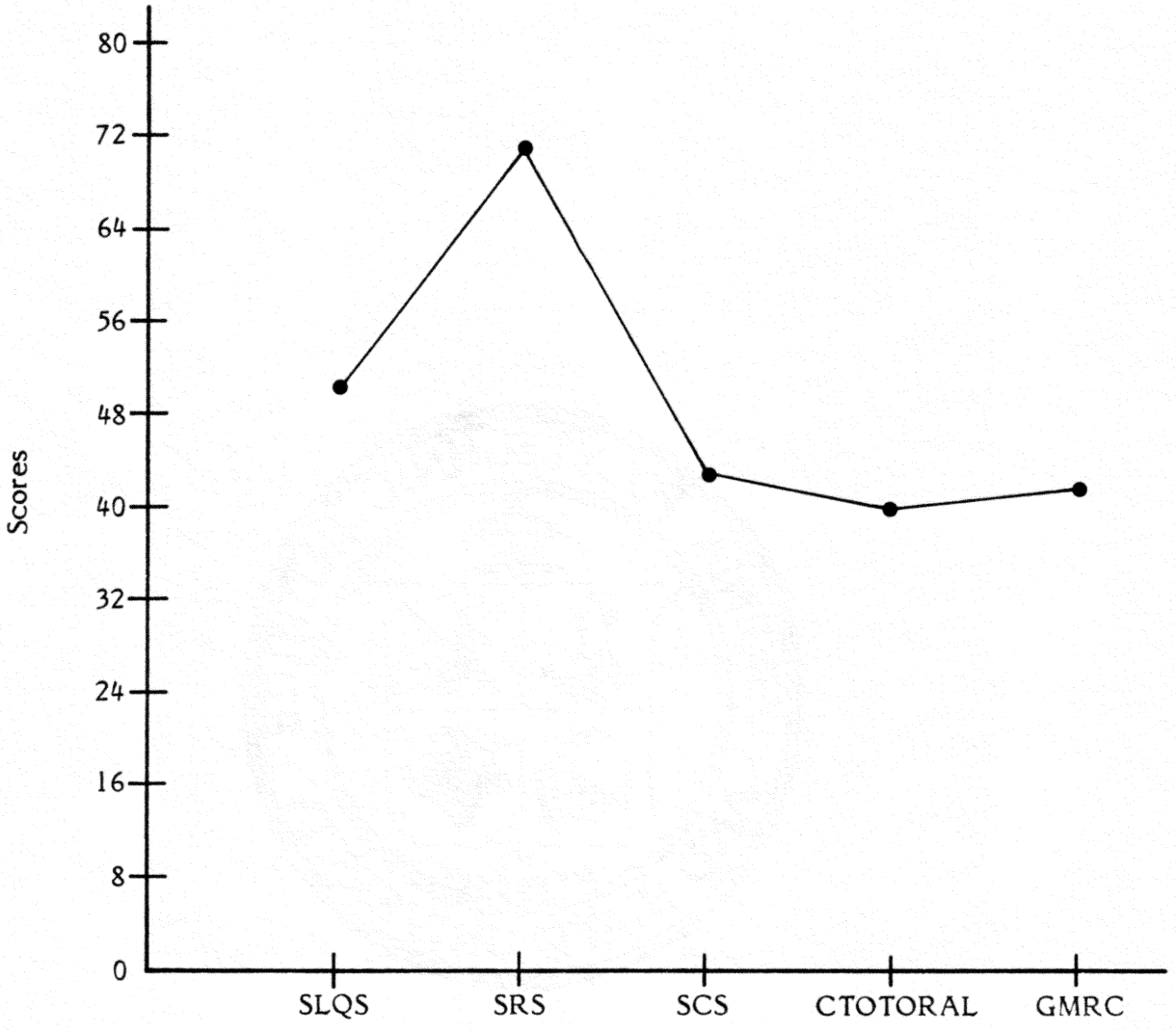
Subject 40's retelling behavior was characterized by frequently saying and/or repeating a word or phrase while the retelling was in process. This seemed to be a way of getting a running start on or rehearsing words chosen for



Alternative Reading Comprehension Assessments

SLQS - Questions Assessment
SRS - Retelling Assessment
SCS - Cloze Assessment
CTOTORAL - Miscue Assessment
GMRC - Standardized Assessment

FIGURE 7. Comprehension assessment profile for subject 40, high on product, low on process measures.



Alternative Reading Comprehension Assessments

SLQS - Questions Assessment
SRS - Retelling Assessment
SCS - Cloze Assessment
CTOTORAL - Miscue Assessment
GMRC - Standardized Assessment

FIGURE 8. Comprehension assessment profile for subject 52, high on product, low on process measures.

the actual retelling itself. Very often, the behavior would then be followed by the retelling of a more complete statement derived from or built upon the stated or repeated phrase, such as, "It was a gorilla . . . a gorilla . . . it was a baby gorilla." In addition, subject 40 demonstrated notable subvocalization while reading the passage prior to retelling.

This characteristic of repeated retelling with elaboration was observed in at least eight other cases in the sample and is somewhat analogous to S. Smith's (1979) description of how "language relay" (pp 91-92) problems may affect retellings. Smith writes that in the process of trying to retell an idea, an individual's thinking may sometimes be diverted by a search for the appropriate language to express that idea. When this happens, it is usually apparent in the language of the retelling itself. The instances of language relay noted in the retellings of subjects in this study, however, tended to operate more like "stalls" to allow for word search to occur rather than as interrupters of thinking. In fact, word search-type pauses and repetitions seemed to be facilitators rather than inhibitors of expression, overall. This was particularly interesting in view of the fact that difficulty in oral expression is frequently listed among the characteristics of learning disabled individuals.

Subject 52's retelling was two standard deviations above the sample mean for the assessment. The striking feature of his protocol dealt with a pronoun referent. Throughout the entire, detailed retelling of a story about a female gorilla who had learned sign language, all references to the gorilla in the retelling were masculine ("he was," "his fingers," etc.). At the conclusion of the retelling, however, the subject's response to the researcher's question about whether or not he remembered anything else about the passage was, "The gorilla's a girl . . . (long pause) . . . that's all." For that reason, the subject was given full credit for all information retold, even though he had used an incorrect pronoun referent throughout the retelling.

On the other product measure, questioning, both subjects answered eight out of ten questions correctly, the same as the mean for the sample. Subject 40, who made the repetitions with subsequent elaboration during retelling, produced four alternative answers (out of his total of eight correct responses) that were not on the prepared answer key. He was given full credit for each of them according to the guidelines for modifying the questioning answer keys described in Chapter Three. Subject 52 exhibited no unusual question answering behavior, but it should be noted that close examination of his assessment profile (Figure 8) shows that his performance on the questioning assessment was more similar to his cloze and miscue performance than to his

retelling performance. For this subject, the high-product/low-process score pattern was not as clear as it was for Subject 40.

For both subjects 40 and 52, cloze and miscue performance was in the first quartile (lowest 25 percent) of scores in each distribution. Just as he had subvocalized on the passage read prior to retelling, subject 40 used finger-pointing during the completion of the cloze exercise. He expressed concern about the mutilated text because he felt that the "sentences made sense to start with," meaning that additional words were not always necessary to create meaningful sentences. Subject 40 completed eight cloze blanks (32 percent of the total) correctly after working twenty minutes on the task, the longest time of any other subject in the sample. Even if synonymous replacements had been counted as correct, his total score would have only exceeded the mean of the sample (12 points) by one point.

Subject 52 completed nine cloze blanks (36 percent of the total) correctly and also took an extraordinarily long time to do the task, 17 minutes. This was the fourth longest time for task completion of all other subjects in the sample. In completing the blank, "His father had to raise _____ the lights that were hanging down from the ceiling," subject 52 commented, "That makes sense without the blank." The tenth-word deletion pattern had required the removal of the text word "all" in this case, but subject

52 appeared dissatisfied that he was required to insert a missing word when it appeared unnecessary to him. This was similar to the sentiment expressed by subject 40. Many of subject 52's cloze replacements made sense, even though they were not exact replacements. In fact, had synonyms been counted as correct responses, his score would have risen from nine to 17 points.

For the miscue assessment, both subjects 40 and 52 miscued frequently and meaning was not retained in approximately one-half of the miscues they produced. Subject 40 had a total score for the miscue assessment of 35 points while subject 52's score was 41 points out of a maximum of 75 points.

Subject 40's miscues were characterized by several instances in which one miscue seemed to be linked to the one following it. For example, in a passage about Harriet Tubman's life career of rescuing slaves during the Civil War, subject 40 read that Harriet did not think she was "being traded family" for Harriet did not think she was "being treated fairly." While graphophonemic similarity was apparent in the linked miscues, an examiner might infer that the word "traded" had resulted from the subject's possible prior knowledge of slave trading practices. Both "traded" and "family" were subsequently corrected, however. The same subject also read, "They group workers insisted" for "The group worked in secret," referring to the operation of

members of the Underground Railroad. In this instance, only "The" was corrected. Interestingly, subject 40 read "Harrimet" for "Harriet" throughout his entire oral reading except on one occasion when he read "Harrimet Tubman" and another when he read "Harrimet Hutton." Another subject, not among the five with the most varied scores, even referred to Harriet Tubman during his retelling as "Harrien Truman." Undoubtedly, intrusions from prior knowledge can be documented in instance after instance. Finally, subject 40 also produced a number of non-words, such as "crevery" for "clever" and "exkution" for "exciting."

Subject 52's miscues also reflected an interesting use of language cues. For example, he produced transformations of text language which retained complete syntactic acceptability but which totally changed the text's intended meaning as in, "That's why her owner couldn't make a lot of money . . ." for "That way, her owner could make a lot of money." In addition, his miscues revealed an effort to construct meaning as in, "Harriet had made escaping adventures" for "Harriet had many exciting adventures." Substituting "escaping" for "exciting" was certainly consistent with passage content in this case. Subject 52 also omitted whole words and word-endings frequently, reading "hard" for "harder," and "north" and "south" for "northern" and "southern." A slight indication of difficulty with pronoun referents, as observed in his retelling, was also observed in the oral reading of "she" for "they."

In summary, the profile of subjects 40 and 52 as illustrated in figures 7 and 8 look relatively similar, but on examining the performance of both of them more closely, subject 52 appeared to be the better overall comprehender of text information based on the results of the cloze and miscue assessments. This difference would not have been apparent from the standardized test score or even from the strictly quantitative indexes of the other four measures.

Patterns of most varied scores compared to the entire sample. Tables 9 and 10 compare the performance of the entire sample and (a) subjects with a Low Product-High Process scoring pattern, N=3, and (b) subjects with a High Product-Low Process scoring pattern, N=2, with respect to the sample means for each comprehension assessment investigated. Figures 9 and 10 present the same information in graphic form except that the means for the sample and for subjects with the most varied scores on all comprehension assessments are expressed as a percentage of the maximum number of points possible for each assessment. The contrast between the two scoring patterns of subjects with the most varied scores and the performance of the sample is especially apparent in Figures 9 and 10 where the shaded areas on each line of the bar graph show the extent to which the most varied subjects' scores differed from the sample's performance.

TABLE 9

COMPARISON OF LOW PRODUCT-HIGH PROCESS MOST VARIED SUBJECTS' (N=3) MEANS ON FOUR ASSESSMENTS TO THE SAMPLE MEANS

Assessment	Sample Mean	Most Varied Subjects' Mean
Questioning	8	6
Retelling	24	15
Cloze	12	15
Miscue	56	69

TABLE 10

COMPARISON OF HIGH-PRODUCT-LOW PROCESS MOST VARIED SUBJECTS' (N=12) MEANS ON FOUR ASSESSMENTS TO THE SAMPLE MEANS

Assessment	Sample Mean	Most Varied Subjects' Mean
Questioning	8	8
Retelling	24	37
Cloze	12	9
Miscue	56	38

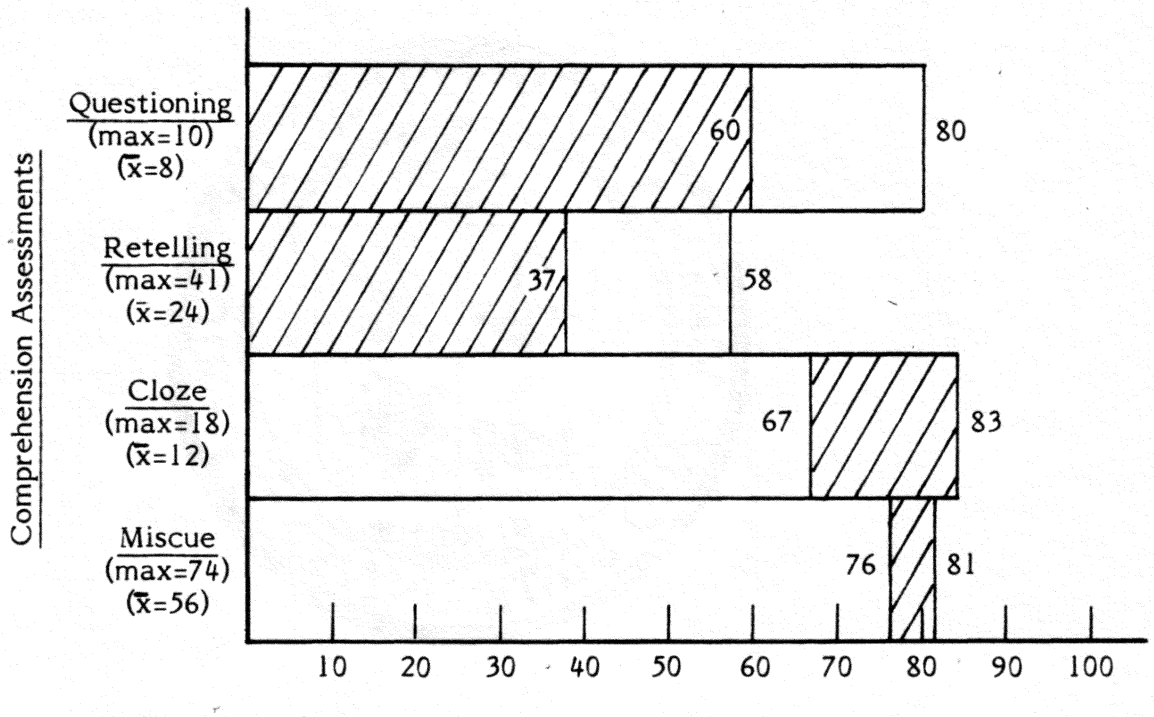


FIGURE 9. Performance of the most varied subjects with Low Product-High Process score pattern (N=3) compared to performance of the sample based on a conversion of the mean for each assessment to a percentage of the maximum score possible.

(Shaded area represents performance of the most varied subjects.)

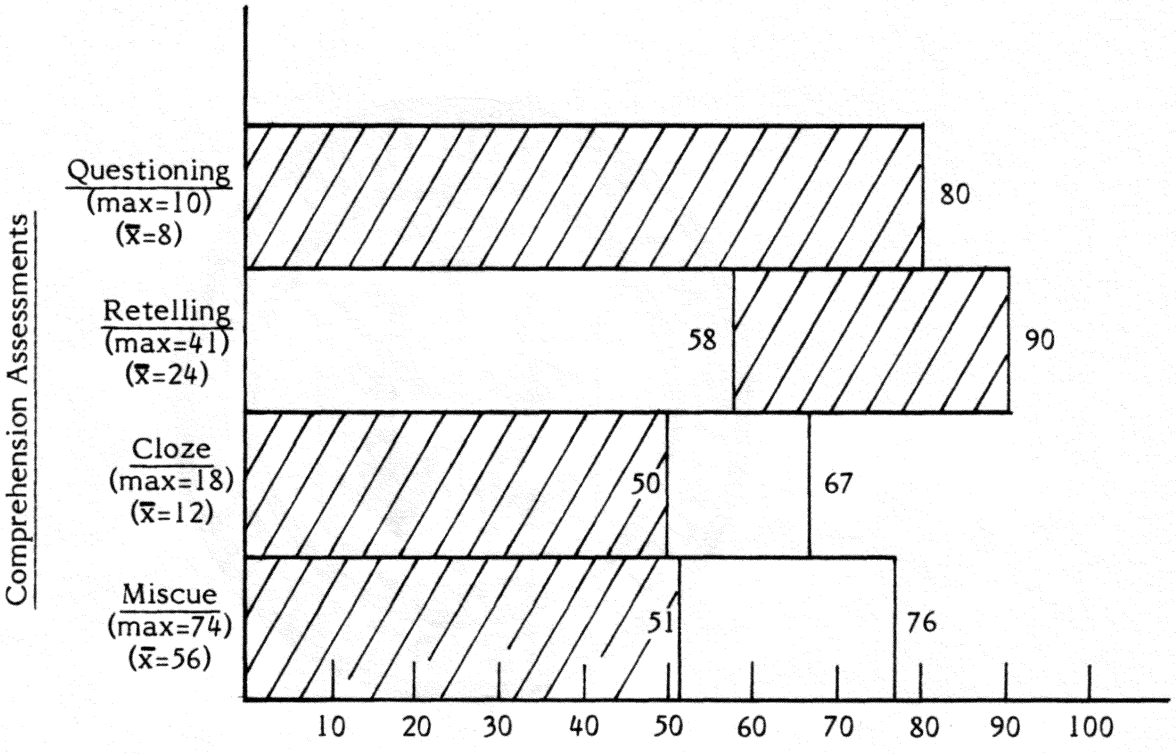


FIGURE 10. Performance of the most varied subjects with High Product-Low Process Score pattern (N=2) compared to performance of the sample based on a conversion of the mean for each assessment to a percentage of the maximum score possible.

(Shaded area represents performance of the most varied subjects.)

Patterns of the most varied scores as related to hypothesized comprehension assessment models. Looking in depth at the comprehension behavior of only five subjects revealed patterns of performance that were not in any way statistically significant or generalizable. Nevertheless, assessment profiles lent support for the assumptions of assessment model two, that the results of alternative comprehension assessments are not interchangeable, and for the assumptions of assessment model three to the extent that subjects' scores were either high on product measures and low on process measures or the reverse. For these five individuals, scores varied according to category of assessment, without exception. Twelve other subjects in the sample had assessment profiles which varied according to assessment category, six for the high product/low process pattern and the same number for the low product/high process pattern, even though their scores were not identified as "varied" according to the criteria explained in Chapter Three. The fact that a total of 15 subjects or 31 percent of the sample exhibited scoring patterns that differed by category of assessment leads one to question whether differences between product and process measures of comprehension assessment were greater than was apparent in the correlational analysis of data. Further research is needed to examine these differences.

In summary, five subjects, comprising 10 percent of the sample studied, showed extreme variability in their scores on the questioning, retelling, cloze, and miscue assessments. For each of the subjects, scores in one category of assessment (product or process) were high while scores in the other category of assessment were low. In addition, reading grade equivalents on a standardized reading test were lower than current grade placement levels for each of the five most variable subjects identified.

Subjective observations. Subjects exhibited a variety of behaviors that illustrated their involvement in trying to construct text meaning. In some cases, these behaviors also revealed their particular difficulties with the comprehension process. Regardless of subjects' strengths and/or weaknesses, each comprehension assessment seemed to generate different response patterns which were loosely categorized as physical mannerisms displayed during testing, approaches to the various assessment tasks themselves, and responses indicating unique interactions with text.

The majority of the physical mannerisms displayed during testing seemed to be reflections of possible anxiety. Some subjects asked that questions be repeated. Others demonstrated restlessness in their behavior such as squirming and rubbing of eyes and face during testing. Several subjects expressed a view that the retelling assessment was hard while others could not see a reason for completing

cloze blanks when they felt that sentences made sense as written without inserting additional words. They seemed to have trouble understanding that the cloze blank held the space for a word that normally appeared in the text. Finally, subvocalization was observed during silent reading and finger-pointing was used occasionally as an apparent aid to smooth tracking across the line of print.

Several commonalities were noted in subjects' approaches to the various assessments tasks themselves. In the questioning assessment, responses to questions were frequently short and lacked richness of vocabulary, information, and sentence structure. Subjects seemed to be searching the hardest for literal and accurate answers, and if the passage did not supply a precise answer to a question asked, then reluctance to formulate an integrated answer on their own was observed. Subjects seemed particularly uncomfortable with the "scriptally implicit" questions which required them to draw on their knowledge/backgrounds for the answer. This was especially the case with questions that started "Why do you think . . ." It was almost as though subjects did not realize that offering an opinion was as viable a response as the recall of factual information. Interestingly, subjects recalled information relating to a number when the number information appeared in a numeral form in the text. When numbers were written in word-form, they were rarely recalled.

Subjects' individual responses were most interesting in the retelling assessment, and a more in-depth study of the retell protocols collected in this study should be undertaken at a later time. When retelling, subjects seemed to have the most trouble sequencing passage events in the proper order. Because of this difficulty, the idea units retold were rarely connected in any logical way. This was particularly evident when a time sequence was important to establishing logical connections between idea units. To some extent, every passage used in this study had a "time sequence" element, but it seemed to be more of a problem for retelling when events in a passage spanned a number of years. When this was the case, particularly in "Harriet Tubman" and in "Gorilla," subjects retold events in a haphazard way, skipping around in the time sequence, with no apparent awareness of what happened first and what happened later. Interestingly, one subject asked to read his passage a second time because he was unsure about the "dates."

As explained in an earlier section of this Supplementary Analysis, subjects frequently "stalled" in their retellings by repeating a certain word or phrase until able to produce another new chunk of information. The observer could almost feel subjects trying to determine what they would say next during this "pause." Following the "pause", a more fluent flow of thoughts was frequently observed. While some retellings contained a considerable amount of

information, a portion of them were extremely limited. It seemed that students' expressive language problems may have been a contributing factor in this restriction. In addition, passages were fairly long and contained much detailed information. Retrieval and memory problems may also have affected the quality of retellings.

In general, subjects also exhibited uncertainty about pronoun referents in their retellings. When subjects completed their retellings and were then asked whether they remembered anything else (followed by a "wait" time), many subjects responded with at least two additional idea units. It was unusual that a subject elected not to contribute additional information at this point.

Subjects did not experience the level of difficulty with the cloze assessment that the researcher anticipated. Basically, the assessment task was transformed by subjects into an oral reading of passages rather than a silent reading as originally designed. It seemed that the process of giving the researcher an oral indication of certain words to be inserted in the cloze blanks tended to make the task more interactive in nature. Likewise, subjects seemed to want to "hear themselves" read aloud as a way of checking whether a particular word choice for a cloze blank "sounded right." Some subjects reread a given sentence or part of a sentence several times, trying out different cloze responses each time, until they were satisfied with their selection. Many

excellent synonomous word choices were produced. Finally, while subjects appeared to use text information prior to a given cloze blank as an aid in completing that blank, many did not appear to read "past" the blank to get additional information from the remaining portion of the sentence. They operated as though text appearing prior to the blank was not related to text which followed the blank. Subjects tended to guess somewhat wildly at cloze answers if they had no ideas for an appropriate response. Some would say they knew their choice was not correct, but they were willing to complete the assessment. While many subjects worked rather slowly on the cloze assessment compared to the other measures, no one refused to complete it out of frustration.

The most interesting aspect of the miscue assessment was observing miscues which subjects produced that were linked to each other as a unit. Some of these miscues are listed in the description of the "most varied cases" section of this chapter. Subjects' oral reading occasionally reflected a relocation of punctuation, usually done meaningfully. Some of the same difficulties with pronoun referents observed on the retelling assessment were observed here. Overall, however, subjects showed that they used a variety of language cues when reading. Even though this LD sample's standardized reading test scores were generally lower in comparison to regular education students, their inherent ability to use a variety of sources of information to get

meaning while reading was evident. This was particularly apparent in the large number of self-corrections produced.

The final striking observation of subjects' performance on all four comprehension assessments was the extent to which they appeared to incorporate whatever background knowledge they had (closest available schema?) into the process of understanding text. Sometimes this aided comprehension and sometimes it contributed to a distortion of meaning. Examples of this phenomenon occurred again and again but the scope of these comments does not permit a detailed review of the diverse illustrative examples.

In observing subjects' performance during data collection, comprehension performance seemed to be adversely affected by memory, retrieval, organization, anxiety, fund of knowledge and experience, and expressive language problems. On the other hand, it seemed to be positively affected by subjects' expectation that text was supposed to make sense, by their incorporation of prior knowledge into the comprehension process, and by their use of the various cueing systems of language to predict text meaning.

Summary of Chapter Four

This chapter dealt with the analyses of data collected from a sample of 48 learning disabled subjects to determine the differences in subjects' performance on four alternative measures of comprehension, to investigate the nature of relationships among the four measures, and to identify which

of the four measures best predicted subjects' standardized test scores. A summary of the findings for hypothesis one and two was presented. In addition, a supplementary analysis of data reviewed the assessment profiles of five subjects whose scores on the alternative indicators of comprehension were determined to be the most varied in the sample. Patterns in the most varied subjects' scores were reviewed. Finally, the researcher's subjective commentary on subjects' behavior during data collection and on their responses to the task demands of the four assessments was provided.

CHAPTER 5

Summary and Conclusions

Restatement of Purpose

This investigation was conducted for the purpose of examining the nature of relationships between and among the results of alternative measures of reading comprehension. In addition, it was conducted to determine the extent to which three possible patterns of relationship among alternative measures, if any, would be reflected in data collected from a sample of identified learning disabled students.

Significance of the Study

The study posed two questions, one concerning the nature of relationships between alternative measures of reading comprehension (questioning, retelling, cloze, and miscue) and the second concerning the relationships observed between these four alternative measures and a standardized test.

To answer the first question, three possible patterns of relationship between alternative comprehension measures were proposed as different models of comprehension assessment. Model I suggested high correlation ($r \geq .70$) among the four alternative measures. Model II proposed little if any correlation among the four alternative measures ($r \leq .30$). Model III postulated a higher average correlation within alternative assessment types than across alternative assessment types.

To answer the first question, a correlational analysis of subjects' scores on alternative comprehension measures was done to determine to what extent relationships were supportive of any of the three models of comprehension assessment.

To answer the second question, multiple correlation procedures were used to determine how similar subjects' scores on the questioning, retelling, cloze and miscue assessments were to their scores on a standardized test.

While a number of researchers have investigated relationships between alternative measures of reading comprehension, this study was different from other studies reported in the literature because it conducted an explicit test of three possible patterns of relationship that might be expected to occur among alternative measures of reading comprehension. The significance of the study was enhanced by the examination of these relationships in test results collected from a sample of identified learning disabled students. The distinguishing characteristics and sometimes subtle learning problems of LD individuals, particularly with respect to the reading process, lent a unique element of interest to the study's findings.

Overall, the primary importance of this study concerned the implications that could be drawn from the data concerning the interchangeability of alternative measures for assessing students' comprehension in the classroom. To

this end, the study generated an empirical basis for focusing teachers' attention on the advantages of using a "corroborative framework" (Page & Vacca, 1979, p. 51) of assessments in evaluating students' comprehension. Operationally, such a framework of different comprehension assessments would give learning disabled individuals a chance to express their understanding of text on alternatives to the standardized test. In the same way, use of different comprehension assessments might encourage teachers to examine differences in reader behavior resulting from alternative indicators of comprehension and to apply knowledge of those differences in planning instruction.

Without a study such as this one and others like it dealing with relationships among various comprehension assessments, it is likely that teachers might continue to be largely uninformed about alternative indicators of comprehension, their similarities, their differences, their strengths, their limitations, and the extent to which each indicator might or might not contribute different information to the profile of a learning disabled student's comprehension performance. Knowledge of assessment alternatives is critical to understanding comprehension, to offering LD students options for expressing their comprehension, and to carrying out instructional planning in reading at any level.

Summary of Procedures

Forty-eight learning disabled males comprised the sample in this study. Subjects were either seventh or eighth graders at one of three intermediate schools in a school district located in metropolitan Washington, D.C. The mean full Scale WISC (Wechsler, 1974) IQ score for the sample was within the average range of intellectual functioning and most subjects' reading grade levels as measured by a standardized test were within the third-to-fifth grade level range.

During the four month data collection phase of the study, subjects read four narrative passages, two each on two successive days. Each passage had been adapted by the researcher from commercially available comprehension practice materials. The overall difficulty level, length, and interest level of each passage were similar.

Comprehension of each passage was assessed by either a questioning, retelling, cloze, or miscue assessment, which had also been developed by the researcher. Questioning and retelling assessments were considered product measures of comprehension, while cloze and miscue assessments were considered process measures of comprehension. The order of administration of test passages and of assessment conditions was systematically varied to avoid serial effects as well as to avoid the confounding of passage and assessment condition. In addition, each subject took a group-administered

standardized test of reading achievement near the end of the data collection period. Data were collected from January to May, 1982.

Summary of the Analysis

To answer the first question posed by the study relating to the nature of relationships among alternative measures of reading comprehension, a correlational analysis was conducted for pairs of scores for each subject, including two product measures of comprehension, two process measures of comprehension, and one standardized measure of comprehension. Pearson Product-Moment coefficients of correlation were organized and displayed in the standard matrix format. All obtained correlations (except those found to be near zero) were tested for statistical significance from zero at a probability level of $< .05$. Correlational data were interpreted by examining the coefficients of determination between pairs of variables.

To answer the second question posed by the study, multiple correlations were utilized in a regression equation to determine which comprehension assessments investigated in the study most closely matched subjects' performance on a standardized measure of comprehension.

Summary of Findings

Hypothesis one: the observed data will reflect one of three possible patterns of relationship among four indicators of comprehension thereby supporting a specific

model of comprehension assessment. Data analysis showed that for this sample of seventh and eighth grade learning disabled boys, no postulated model of comprehension assessment was completely supported by the relationships observed.

First, the data were entirely inconsistent with assessment model I, there being no correlation among informal measures at $\geq .50$.

Second, the data provided limited support for assessment model II, low relationships among alternative measures of comprehension. Support for model II was considered limited because three of the six possible relationships observed in the data (questioning-retelling; retelling-cloze; retelling-miscue) were within the hypothesized range of $\leq .30$, but the remaining three of the six possible relationships observed in the data (questioning-cloze; questioning-miscue; miscue-cloze) were greater than the hypothesized criterion for low relationships. Nonetheless, none of the three relationships greater than .30 exceeded a magnitude of .50. For this reason, correlations were considered to be representative of a low association among measures overall. In addition, an average of the six possible relationships observed in the data was .26, well within the hypothesized range describing a low level of association among variables.

Third, the data provided limited support for assessment model III, higher relationships within assessment type

(product, process) than across assessment types. Although an average of correlations within the product and within the process assessment types was higher than an average of the across-type correlations (.38 compared to .20), thus confirming a slight distinction between product and process measures of comprehension, the distinction was diluted by the observation that the highest correlation among alternative indicators of comprehension in the study crossed assessment categories (in the relationship between questioning/product and cloze/process measures). This finding suggests that the questioning (product) task and the cloze (process) task may be measuring some of the same factors, despite their description as different types of assessment. Moreover, the low across-type average of correlations among measures (r=.20) resulted from the fact that virtually any correlation which involved the retelling measure was either low or nonexistent in this study. When correlations involving retelling were removed from the calculation of the across-type average, the resulting r was .39, almost the same as the within-type average of .38.

Hypothesis two: relationships of alternative measures of comprehension to a standardized test. Regression analysis procedures were applied to determine which of the four alternative comprehension assessments (questioning, retelling, cloze, and miscue) best predicted subjects' performance on a standardized test.

Data analysis showed that subjects' performance on the miscue, cloze, and retelling assessments (in the order of their contribution to the multiple R) accounted for nearly 50 percent of the variance associated with variance on the standardized measure of comprehension.

Discussion

Comprehension is still not directly observable, regardless of the measurement strategy one uses. It is also a complex process, apparently affected by a myriad of factors, many of which remain unknown. But observing reading behavior is the only option available for decreasing our ignorance of the phenomenon.

In this study, there were many opportunities to speculate about comprehension by considering whether different comprehension assessments, formal and informal, really measured different aspects of comprehension and whether they might be interchangeable (with each other) in practical use, particularly in evaluating the comprehension of learning disabled students. To this end, a correlational analysis of the relationships observed among and between informal assessments and a standardized test provided a basis for evaluating how subjects' scores varied under different testing conditions.

The analysis of hypothesis one was interesting for the sample of LD students tested in this study for two reasons. First, it would appear from these data that the

questioning, retelling, cloze and miscue assessments were largely exclusive of each other as observed in the low relationships among measures, expressed most closely by assessment model II. Although not strong, assessment model II was the only one of the three proposed models of comprehension assessment that was most nearly supported by the data. On the basis of a pattern of low intercorrelations among measures as suggested by assessment model II, the questioning, retelling, cloze and miscue measures investigated in this study cannot be viewed as totally interchangeable for assessing reading comprehension. However, the data actually failed to confirm any of the three models of comprehension assessment. Rather, they were more consistent with a model of assessment viewing retelling as one cluster of comprehension measures and questioning, cloze, and miscue as another cluster of comprehension measures.

Second, the data suggested that distinguishing between measures by category of assessment (product, process) was not as important to measuring comprehension for this group of LD students as were the differences observed between each alternative method of assessment. Methods of assessment used in this study were somewhat different in underlying theory and in task demands. These differences were particularly noticeable for the retelling assessment which produced more variability in subjects' responses than did any of the other three alternative assessments. Clearly,

the retelling task seemed to require subjects to engage in comprehending behaviors that were not the same as those required by the questioning, cloze and miscue assessments, and which probably require greater effort in the construction of a response.

The analysis of hypothesis two was interesting for the sample of LD students tested in this study because even though the multiple R in the regression of alternative assessments on the standardized test was high (.71), the additive effects of several alternative assessments (each different from the other in theoretical bases, assessment task, and association with other measures) were required to account for only 50 percent of the variance associated with the standardized test. This observation tends to verify a multiple-measures approach to comprehension assessment.

Even though this shared variance among alternative measures and the standardized test was limited, it is reasonable to conclude that especially the miscue and cloze assessments (entered at step one and two of the regression equation) and the standardized test seem to measure many of the same factors involved in comprehending. But the analysis also suggests that as much or more credence should be given to the results of miscue and cloze assessments in measuring comprehension (as compared to the standardized test) because miscue and cloze have stronger theoretical bases in explaining the reading process and the standardized

test has no such basis. Further, an assessment such as miscue will undoubtedly identify many more areas where instruction is needed, especially in the more effective use of certain cue systems, than will be evident in standardized test results.

A more precise identification of the factors associated with the unaccounted-for variance between the assessments identified in the regression equation and the standardized test requires further investigation. This variance could be the result of a multiplicity of factors including such variables as subjects' expressive language skill, visual format of a comprehension assessment (particularly applicable to cloze), size of text print, length of passage, subjects' distractibility, ability to follow oral and written directions, time pressures, anxiety of the test situation, willingness to risk, self-motivation, and so forth.

By the overall trend in the data toward a pattern of low relationships observed among alternative comprehension assessments, this study supports the assertion that comprehension is a cognitive ability which is more complex than it is unitary. Further, some of the many dimensions of comprehension associated theoretically with the four alternative assessments examined appear to be tapped to some extent by all of them, but not to the extent that one assessment can effectively replace another assessment in evaluating students' abilities. In addition, some moderate

relationships among measures (particularly questioning-cloze; cloze-miscue) suggest that the results of multiple comprehension assessments may corroborate each other somewhat, except when the task requirements of a particular assessment may cause an examiner to anticipate wide variations in a reader's expression of comprehension, as was the case with the retelling assessment in this study. An awareness of potential difficulties in expressing comprehension on certain assessments is especially important in working with identified handicapped students whose learning problems may preclude success with particular assessment tasks.

In brief, differences in assessment task requirements of the various comprehension measures examined in this study coupled with the low relationships observed among these measures in the data tend to confirm the assumption that a reader's comprehension output may be partially a function of how it is measured. This assumption has important implications for practice.

Implications for Education

Data in this study were gathered and analyzed not only to determine what could be learned about how students comprehend text when alternative comprehension assessments are used, but also to provide a basis for speculating about how instructional options might be broadened when information about comprehension is collected from alternative sources.

This section addresses those instructional options resulting from comprehension assessment.

When a teacher encounters a new group of students, the first comprehension measure given is usually a standardized test of reading achievement. This is done to obtain a current level of achievement for each student in the group and is a routine procedure in special education because objective test data are considered part of every student's Individualized Education Program (IEP).

While a standardized test is relatively easy to administer and to score, its results give the teacher almost no useful information for making instructional modifications in teaching reading. Again, in special education programs, a teacher expects to make many such modifications because the varied abilities and skill levels of identified handicapped students require them. But modifications should not be made for teaching reading comprehension unless a teacher collects and analyzes as much information about the current state of a reader's comprehending abilities as possible. What information does this study provide to suggest how teachers can accomplish this goal most effectively?

The data in this study suggest that four alternative indicators of reading comprehension (questioning, retelling, cloze, miscue) have some utility for assessing comprehension and that these measures cannot be effectively substituted

for each other in an assessment repertoire because each one contributes slightly different information about a reader's understanding of text (which is partly a function of different assessment task demands). In fact, this study confirms that only when a group of alternative measures is used in assessing comprehension is there likely to be a higher association with variance in standardized test scores. Plainly, assessing comprehension with more than one alternative to the standardized test would seem to be sound educational practice with handicapped students.

How should teachers implement a multiple-measures approach to assessing comprehension? One obvious way would be to give each alternative assessment to all students and then compare their performance across measures. Such an approach has obvious limitations, the major one being the time required to individually administer and score so many assessments on every class member.

To save time, a teacher might try a second approach to assessment which would involve giving each alternative measure to only those students who produced low scores on a standardized reading test. With learning disabled students, however, this approach could require a nearly equal expenditure of time as the first option because many of them are likely to produce low scores on the standardized test. Moreover, there are students, the data suggest, who make high scores on the standardized test but who are also weak

in abilities not clearly measured by it but possibly critical to expressing reading comprehension, such as monitoring one's own search for and retention of meaning.

A third approach to assessment might be to supplement information from the required standardized test administration with a selection of alternative comprehension assessments (but not all of them) and only with selected students. If the results of this study had strongly supported assessment model III, this approach would be a good alternative; hypothetically, the teacher could give one assessment from a given category of measurement (product) along with another assessment from a second category of measurement (process) in order to assess a wider range of comprehending behaviors exhibited by students. Since the data did not strongly support a distinction between measures contained in product and process assessment categories, however, this approach has questionable validity. Nevertheless, differences in assessment tasks among measures examined in this study suggest that teachers might choose assessments on the basis of the kinds of instructional decisions they need to make and that certain assessments might be more applicable to identifying the need for instruction in certain areas (e.g., use of syntax, selective cue use, logical connections between ideas) than others.

Despite this limited support for differences in product and process assessments, some subjects in the present study

made high scores on measures in one category of assessment and low scores on measures in the other category of assessment. What this finding may really suggest is that use of a multiple-measures approach to assessment may be one way to respond to the fact that a reader's comprehension output may be partly a function of how it is measured. The viability of comprehension assessment categories is an area where additional research is needed. Indeed, there are undoubtedly comprehension assessments which have yet to be developed and categories of assessment which have yet to be defined.

Since each of the preceding approaches to comprehension assessment with multiple-measures may be problematic, perhaps the most effective and efficient way to accomplish the task in the classroom would be to adapt assessment strategies for instructional purposes. This approach would be feasible because the methodology associated with alternative comprehension assessments can be transformed easily into teaching strategies. For example, a questioning strategy can be used during instruction to help focus students' attention on important text information; cloze encourages an interaction between reader and text through the manipulation of syntactic structures; retelling helps students organize their thinking about text and construct logical connections among ideas; and an analysis of one's own miscues helps students monitor their retention of meaning through the self-correction process.

A special education teacher would derive an additional benefit from the instructional application of comprehension assessment techniques because as teaching progressed, the simultaneous evaluation of comprehension skills would be facilitated through the interchangeability of teaching and testing strategies. This capability fulfills another requirement of the IEP which calls for continuous, frequent evaluation of student progress toward specified goals, and continuous evaluation can only really be accomplished in the context of daily instruction.

Before using alternative assessments to make instructional decisions in the classroom or as modified instructional techniques themselves, it would be ideal if teachers understood the major definitions of comprehension, the theoretical models of the reading process on which each definition was based, and the appropriate assessment/instructional strategies which logically flowed from each model. It is not always apparent that teachers have such understandings, however, nor can they be developed over the short term.

Progress can be made toward increasing teacher awareness through a coordinated series of inservice activities where professionals would learn what each comprehension assessment attempted to measure, how it related to other assessments, and how its limitations affected data collection in general. The overall objective of such an

inservice series would not be to have teachers make a commitment to using all four comprehension assessment alternatives with every student on a regular basis. Rather, it would be to help teachers recognize that if they do give only one comprehension assessment, they might get a different picture of what a student comprehends when they give another assessment. Such a recognition would have great potential benefit for the assessment of learning disabled students' comprehension. These students are likely to need more than one chance to express their comprehension because of the inherent limitations to expression that result from their handicapping condition. Recognizing that comprehension responses may vary by assessment would have great potential for improving reading instruction in the broad sense, and for narrowing the ever-present gap between theory and practice.

In view of the complexity of comprehension itself, the measures of relationship observed in this study reinforced the view that product as well as process assessments seem to be supplying some different and varied information about how a student comprehends and should be employed in the classroom. In the same way, product and process instructional strategies may be useful for developing greater student competency in getting meaning from text.

Suggestions for Further Research

Measuring readers' comprehension with alternative assessment strategies should be an integral part of the research that continues to seek an understanding of the reading process. Researchers and practitioners need to be in direct and frequent contact with readers who are demonstrating their comprehension in alternate ways. This is the only way to identify observable patterns in comprehending behavior, to find out whether those patterns corroborate what theoretical models of reading suggest may be important aspects of comprehension, and to generate hypotheses about what other factors may be affecting readers' comprehension, both negatively and positively.

Suggestions for further research are described in this section. Some of the suggestions result from the methodology and findings of the present study while other suggestions reflect those typically offered in the literature on comprehension assessment.

Further research suggested by the present study. Data analysis in this study has identified a number of areas for further research in assessing comprehension.

First, comparing the performance of a sample of LD students to a sample of regular class students on the same alternative indicators of comprehension as were examined in this study would be useful in describing differences in the comprehending abilities of identified-handicapped and

non-handicapped individuals. Such a comparison would require equating both groups on variables such as age, sex, intelligence, grade, and reading grade level. If patterns in the relationships among alternative comprehension assessments were consistent across both categories of students, one's handicapping condition might not appear to be as significant a source of variability in comprehension performance as factors associated with the assessment tasks themselves. Such a comparison might also lead to additional questions about developmental differences in the two groups, differences in the groups' expectation for meaning in text, their use of language cues in reading, their logical organization of ideas for retelling, their application of prior knowledge in comprehending, and so forth.

Second, redesigning specific assessments used in this study to measure more appropriately aspects of comprehension that research now concludes are important to the total reading process (such as reader background, memory, purpose for reading, interest, inferencing ability) would be useful in increasing the validity of measures. In such a study, the questioning assessment could be constructed to generate different scores for student responses to textually explicit, textually implicit, and scriptally implicit questions instead of providing a composite score for the total assessment. In so doing, it would then be possible to observe how patterns of relationship among measures change

in a correlational analysis of scores when comprehending behaviors such as inferencing and using prior knowledge are specifically accounted for in subjects' answers to questions. Would textually explicit or textually implicit scores correlate more closely with the results of a retelling, cloze, or miscue assessment? What would be the relationship observed between scriptally implicit scores and retelling scores or scores from other assessments?

The retelling assessment could also be redesigned to account specifically for inferential and prior knowledge information in scoring retelling protocols. The cloze measure could examine differences in subjects' responses when variations on the cloze task were incorporated into assessment. Likewise, the miscue assessment could generate scores based on any combination of a reader's use of graphophonic, syntactic, or semantic cues as well as scores representing a reader's monitoring of his/her own meaning through the self-correction process.

The analysis of data in this study found low associations among the results of alternative indicators of comprehension. But if existing assessments are refined so that their task demands are more representative of the major comprehending behaviors identified in the literature, would relationships among measures change? Would relationships suggest that certain assessments might be eliminated from a diagnostic repertoire because of their similarity to other

assessments in type of comprehending behavior measured? On the other hand, would relationships suggest that certain other assessments might always need to be part of an overall comprehension assessment routine because they seem to tap comprehending behaviors not revealed by other measures? How closely would the results of redesigned or newly developed comprehension assessments match students' performance on the standardized test? Analyses of scores from new indicators of comprehension might result in new patterns of relationship among measures that have not yet been observed. This would, in turn, generate additional hypotheses for further study.

Third, continuing to look more closely at alternative assessment strategies as product and process measures of comprehension would be useful. While data in this study did not support a strong distinction between these two categories of comprehension, slight evidence of differences between them in subjects' responses was apparent, and the supplementary analysis of the most varied score profiles in the sample suggested that subjects' comprehension tended to vary along product and process lines. If existing comprehension asssessments were redesigned more nearly to tap behaviors that theoretical models of reading suggest are crucial to comprehending, or if completely new comprehension assessments were developed on the same basis, would these measures fit into a product category (reproducing exact text

meaning), into a process category (integrating language use, reasoning, and existing knowledge to get meaning), or into a new, and as yet undescribed category of assessment?

Fourth, conducting an explicit study of the impact of classroom instruction on the measurement of comprehension would be helpful in establishing a closer link between the testing and teaching process and between theory and practice. In fact, if assessment is done to identify and to clarify instructional goals, how would the patterns of relationship among alternative indicators of comprehension change if teachers used the same strategies to develop comprehension in classrooms that are also used for assessing comprehension? Would higher correlations among measures then be considered a function of instruction and practice and not a function of the aspects of comprehension being tapped by the different assessments? How would the discovery of low relationships among measures be interpreted then?

Finally, this study proposed patterns of association among alternative indicators of comprehension and equated those patterns to specific models of comprehension assessment. In connection with examining testing/teaching procedures in the classroom, further research might be conducted to determine whether teachers exhibited an implicit model of comprehension assessment in their daily routine. Such a study would be concerned with determining whether teachers

(a) tend to rely solely on standardized test data to supply diagnostic information on comprehension, (b) use alternative measurement strategies as part of a frequent and continuous assessment process, or (c) use one category of measurement (product or process) to the exclusion of another.

Further research suggested by the literature. In conducting research with learning disabled students as subjects, the special education literature has supported identification of various learning styles of youngsters (auditory, visual, kinesthetic, tactile, multisensory, and so forth) as a prerequisite to planning instruction to meet diverse learning needs. The reading literature has suggested that readers may also exhibit "comprehension styles" (Spiro, 1980), defined as individual differences in one's preferred way of getting meaning from text. Theoretically, identifying a reader's comprehension style would be the first step in adjusting instruction to meet students' differences in the same way that identifying learning styles would. If further research could document the existence of comprehension styles in learning disabled students, the implications for teaching and testing this group of learners might be significant. Such research would address whether exhibited comprehension styles might in some way be equated to preferred learning styles, and would compare similarities and differences in the observed comprehension styles of LD and regular class students. In so doing, it

might also investigate whether developmental differences in the ability of learning disabled students to comprehend complex text structures might be a factor in comprehension style development.

Reaching a definition of the comprehension process that can be widely accepted by reading experts and teachers continues to be the most important prerequisite for identifying and employing a consistent comprehension assessment methodology. Through the years, the lack of a consistent definition has severely limited the assessment of what is, at the outset, a complex and hard-to-observe cognitive process.

Further research that contributes knowledge to describing and operationalizing the interactions of the cognitive and linguistic components of comprehending will establish a direction for developing new comprehension assessments. The body of existing research shows that advances have been made toward this objective, but that developing comprehension assessments which are more reflective of evolving descriptions of the reading progress should continue to be a high priority.

REFERENCES

Adams, A., Carine, D. & Gersten, R. Instructional strategies for studying content area texts in the intermediate grades. Reading research quarterly, 1982, XVIII, 27-55.

American Psychological Association, Publication manual of the American Psychological Association. 2nd Ed., Washington, D.C.: 1974.

Anderson, R.C. & Spiro, R.J. Schooling and the acquisition of knowledge. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Baldauf, R.B., Dawson, R.L.T., Prior, J., & Propst, I.K. Can matching cloze be used with secondary ESL pupils. Journal of reading, 1980, 23, 435-440.

Balow, B. Perceptual activities in the treatment of severe reading disability. The reading teacher, 1971, 24, 513-525.

Bamman, H.A., Poe, A.M., Raske, D.E., & Thomas, T.A. Breakaway: the EMC basic comprehension series. St. Paul, Minnesota: EMC Corporation, 1980.

Barbe, W., & Abbott, J. Personalized reading instruction. West Nyack, N.Y.: Parker Publishing Co., 1965.

Beebe, M. J. The effect of different types of substitution miscues on reading. Reading research quarterly, 1980, 15, 324-336.

Berg, P.C. (Consultant). Spotlight: scope reading skills 3. New York: Scholastic Book Services, 1973.

Blumenthal, S.H. A study of the relationship between speed of retrieval of verbal information and patterns of oral reading errors. Journal of learning disabilities, 1980, 13, 42-44.

Borg, W.R., & Gall, M.D. Educational research: an introduction. (3rd ed.). New York: Longman, 1979.

Bormuth, J.R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation. Indiana University, 1962.

Bormuth, J.R. An operational definition of comprehension instruction. In K.S. Goodman & J.T. Fleming (Eds.), Psycholinguistics and the teaching of reading. Newark, Delaware: International Reading Association, 1969.

Bormuth, J.R. Factor validity of cloze tests as measures of reading comprehension ability. Reading research quarterly, 1969, IV, 358-365.

Brown, S.M. Skillbooster series. Cleveland, Ohio: Modern Curriculum Press, 1975.

Bryan, T.H. Learning disabilities: a new stereotype. Journal of learning disabilities, 1974, 7, 304-309.

Calfee, R.C., Drum, P.A., & Arnold, R.D. What research can tell the reading teacher about assessments. In S.J. Samuels (Ed.), What research has to say about reading instruction. Newark, Delaware: International Reading Association, 1978.

Carey, R.F. A psycholinguistic analysis of the effects of the semantic acceptability of oral reading miscues on reading comprehension. Unpublished doctoral dissertation. University of Connecticut, 1978.

Carey, R.F. Cloze encounters of a different kind. Reading horizons, 1979, 19, 228-231.

Carey, R.F. A psycholinguistic analysis of the effects of semantic acceptability of oral reading miscues on reading comprehension. Paper presented at the annual convention of the International Reading Association, Atlanta, April, 1979.

Carroll, J.B. Defining language comprehension: some speculations. In R.O. Freedle & J.B. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V. H. Winston and Sons, 1972.

Carroll, J.B. Developmental parameters of reading comprehension. In J.T. Guthrie (Ed.), Cognition, curriculum, and comprehension. Newark, Delaware: International Reading Association, 1977.

Certley, A.S. Learning disabilities versus reading disabilities: a vexing problem. In C.M. McCullough (Ed.), Inchworm, inchworm: persistent problems in reading education. Newark, Delaware: International Reading Association, 1980.

Chall, J. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Chall, J. A decade of research on reading and learning disabilities. In S.J. Samuels (Ed.), What research has to say about reading instruction. Newark, Delaware, International Reading Association, 1978.

Davis, F.B. Fundamental factors of comprehension of reading. Psychometrika, 1944, 9, 185-197.

Davis, F.B. Research in comprehension in reading. Reading research quarterly, 1968, 4, 499-545.

Davis, F.B. Psychometric research on comprehension in reading. Reading research quarterly, 1972, 7, 628-678.

Deck, D. Validity of word deletion items as a measure of reading comprehension. Paper presented at the annual meeting of the American Educational Research Association, April, 1977.

Deno, S.L., Mirkin, P.K., & Chaing, B. Identifying valid measures of reading. Exceptional children, 1982, 49, 36-45.

Dunn, B.R., Matthews, S.R., & Bieger, G. Individual differences in the recall of lower-level textual information. (Technical Report No. 150). Urbana-Champaign: Center for the Study of Reading, University of Illinois, December, 1979.

Durkin, D. What classroom observations reveal about reading comprehension instruction. Reading research quarterly, 1978-79, 14, 481-533.

Durkin, D. What is the value of the new interest in reading comprehension? (Technical Report No. 19). Urbana-Champaign, Illiniois: Center for the Study of Reading, University of Illinois, November 1980.

Durrell, D.D. Durrell analysis of reading difficulty. New York: Harcourt, Brace and World, 1955.

Elkins, J., Andrews, W.C., Apelt, W.C., Cochrane, K.J., Atkinson, J.K. Multivariate relationship between cognitive and reading measures in third grade children. The exceptional child, 1977, 24, 65-74.

Farr, R. The fallacies of testing. Paper presented at the conference on Reading and National Interest, Bloomington, Indiana, March, 1970.

Farr, R., & Anatasiow, N. Tests of reading readiness and achievement: a review and evaluation. Newark, Delaware: International Reading Association, 1969.

Farr, R. & Tone, B. Text analysis and validated modeling of the reading process (1973-1981): implications for reading assessment. Unpublished manuscript, Indiana University, 1982.

Fitzgerald, T.P., & Fitzgerald, E.F. A cross cultural study of comprehension at the primary and intermediate levels. Educational research quarterly, 1978, 3, 84-92.

Flesch, R.F. A new readability yardstick. Journal of applied psychology, 1948, 32, 221-233.

Fry, E.B. A readability formula that saves time. Journal of reading, 1968, 11, 513-516.

Gates, A.I., & Macginitie, W.H. Survey D reading test: speed, accuracy, and comprehension, grades 4-6. New York: Teachers College, Columbia University, 1965.

Gillespie-Silver, P. Teaching reading to children with special needs. Columbus: Charles E. Merrill, 1979.

Golinkoff, R.M. A comparison of reading comprehension processes in good and poor comprehension. Reading research quarterly, 1975-76, 4, 623-659.

Goodman, K.S. Behind the eye: What happens in reading. In K.S. Goodman & O.S. Niles (Eds.) Reading process and program. Urbana, Illinois: National Council of Teachers of English, 1970.

Goodman, K.S. A psycholinguistic game. In H. Singer and R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Goodman, K. & Burke, C. Theoretically based studies of patterns of miscues in oral reading performance: final report. Washington, D.C.: Department of Health, Education & Welfare, 1973.

Goodman, K.S. & Goodman, Y.M. Learning about psycholinguistic processes by analyzing oral reading. Howard educational review, 1977, 47, 317-333.

Goodman, Y. & Burke, C.L. Reading miscue inventory. New York: Macmillan, 1972.

Gough, P.B. One second of reading. In H. Singer & R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Gould, S.M. Cloze restoration, recall, and discourse processing in middle grades readers. In J.C. Harste & R.F. Carey (Eds.), Monograph in language and reading studies; new perspectives on comprehension. Bloomington: Indiana University, 1979.

Gray, W. On their own in reading. Chicago: Scott, Foresman and Company, 1960.

Gredler, G.H. Learning disabilities and reading disorders: a current assessment. Psychology in the schools, 1978, 15, 226-237.

Grimes, J.E. The thread of discourse. The Hague, Netherlands: Morton and Co., 1975.

Gunning, R. The technique of clear writing. New York: McGraw-Hill, 1952.

Guszak, F.J. Teacher questioning and reading. The reading teacher, 1967, 21, 227-34.

Guszak, F.J. Questioning strategies of elementary teachers in relation to comprehension. In L.A. Harris and C.B. Smith (Eds.), Individualizing reading instruction: a reader. New York: Holt, Rinehart and Winston, 1969.

Guszak, F.J. Diagnostic reading instruction in the elementary school. New York: Harper & Row, 1972.

Guthrie, J.T. Models of reading and reading disability. Journal of educational psychology, 1973, 65, 9-18.

Guthrie, J.T. The 1970's comprehension research. The Reading Teacher, 1980, 33, 880-882.

Hammill, D.D. and Larsen, S.C. Relationship of selected auditory perceptual skills and reading ability. Journal of learning disabilities, 1974, 7, 429-436.

Hansen, C.L. Story retelling used with average and learning disabled readers as a measure of reading comprehension. Learning disability quarterly, 1978, 1, 62-69.

Hargis, C.H. A comparison of retarded and nonretarded children on the ability to use context in reading. American journal of mental deficiency, 1972, 76, 726-728.

Harris, A.J. An overview of reading disabilities and learning disabilities in the U.S. The reading teacher, 1980, 33, 420-424.

Harris, L.A. & Smith, C.B. Reading instruction: diagnostic teaching in the classroom. New York: Holt, Rinehart, and Winston, 1976.

Harste, J.C. & Burke, C.L. Toward a socio-psycholinguistic model of reading comprehension. Viewpoints in teaching and learning, 1978, 54, 9-34.

Heerman, C.E. & Treadway, K.A. Cloze for monitoring and enhancing reading comprehension. Viewpoints in teaching and learning, 1978, 54, 82-89.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. Applied statistics for the behavioral sciences. Chicago: Rand McNally College Publishing Co., 1979.

Hollander, S. K. Reading: process and product. The reading teacher, 1975, 28, 550-554.

Hood, J., & Kendall, J.R. Investigating the relationship between comprehension and word recognition: Oral reading analysis of children with comprehension or word recognition disabilities. Journal of reading behavior, 1979, 11, 41-48.

Hood, J. Qualitative analysis of oral reading errors: the inter-judge reliability of scores. Reading research quarterly, 1975-76, XI, 577-598.

Hutson, B.A. & Gove, M. Reading strategies profile: miscue analysis without tears. Paper presented at the annual convention of the International Reading Association, St. Louis, May, 1980.

James, P.C. The use of cloze with educationally sub-normal children. Reading, 1975, 9, 26-32.

Johnston, P. H. Implications of basic research for the assessment of reading comprehension. (Technical Report No. 206). Urbana-Champaign: Center for the Study of Reading, University of Illinois, 1981.

Johnston, P.H. Reading comprehension assessment: a cognitive basis. Newark, Delaware: International Reading Association, 1983.

Jones, L. An interactive view of reading: implications for the classroom. The reading teacher, 1982, 7, 772-777.

Kaluger, G. & Colson, C.J. Reading and learning disabilities. Columbus, Ohio: Charles E. Merrill, 1978.

Kamil, M.L. Models of reading: what are the implications for instruction in comprehension? S. Pflaum-Connor (Ed.), Aspects of reading education. Berkeley, California: McCutchen Publishing Corp., 1978.

Kaminsky, S. Taking a closer look at the cloze procedure. Reading world, 1979, 19. 12-18.

Karlin R. Evaluation for diagnostic teaching. In W. MacGinitie (Ed.), Assessment problems in reading. Newark, Delaware: International Reading Association, 1973.

Kemp, D.M. Measuring reading performance by miscue analysis. Education news, 1976, 82-87 (reprint).

Kephart, N.C. The slow learner in the classroom. Columbus: Charles E. Merrill, 1971.

Kingore, B.W., & Kurth, R.J. A workshop for teachers in reading comprehension. The reading teacher, 1981, 35, 173-179.

LaBerge, D. & Samuels, S.J. Toward a theory of automatic information processing in reading. In H. Singer & R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Lerner, J.W. Children with learning disabilities: theories, diagnosis, and teaching strategies. 2nd ed. Boston: Houghton-Mifflin Co., 1976.

Lovitt, T.C. & Hansen, C.L. The relationship between question type and mode of reading on the ability to comprehend. The journal of special education, 1976, 10, 53-60.

MacGinitie, W.H. Gates-MacGinitie reading tests. Boston: Houghton-Mifflin Company, 1978.

Marr, M.B., & Gormley, K. Children's recall of familiar and unfamiliar text. Reading research quarterly, 1982, XVIII, 89-104.

Marshall, N. and Glock, M.D. Comprehension of connected discourse: a study into the relationships between the structure of text and information recalled. Reading research quarterly, 1978-79, XIV, 10-56.

Meyer, B.J.F. Organization of prose and its effects on recall. Amsterdam: North Holladn Publishing Co., 1975.

Meyer, B.J.F. The structure of prose: effects on learning and memory and implications for educational practice. In R. Anderson and R. Spiro (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum Associates, 1977.

Miller, J.A. Strategies for diagnosing and developing comprehension skills. Viewpoints in teaching and learning, 1978, 54, 97-106.

Mitchell, C.A. Psycholinguistic analysis of oral reading errors of children with learning disabilities. Paper presented at the meeting of the International Scientific Conference of IFLD, Montreal, 1976.

Moe, A.J. Cohesion, coherence, and the comprehension of text. Journal of reading, 1979, 23, 16-20.

Newcomer, P. & Magee, P. The performance of learning (reading) disabled children on a test of spoken language. The reading teacher, 30, 1977, 896-900.

Newman, H. Oral reading miscue analysis is good but not complete. The reading teacher, 1978, 32, 883-885.

Otto, W. Design for developing comprehension skills. In J.T. Guthrie (Ed.), Cognition, curriculum, and comprehension. Newark, Delaware: International Reading Association, 1977.

Page, W.D. The post oral reading cloze test: New link between oral reading and comprehension. Journal of reading behavior, Winter, 1975.

Page, W.D. Comprehending and cloze performance. Reading world, 1977, 17, 17-21.

Page, W.D. & Pinnell, G.S. Teaching reading comprehension. Urbana, Illinois: National Council of Teachers of English, 1979.

Page, W.D., & Vacca, R.T. Overt indications of reading comprehension: product and process considerations. In J.D. Harste and R.F. Carey (Eds.), Monograph in language and reading studies: new perspectives on comprehension. Bloomington: Indiana University, 1979.

Pearson, P.D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading research quarterly, 1974-75, 10, 155-192.

Pearson, P.D., & Johnson, D. Teaching reading comprehension. New York: Holt, Rinehart, & Winston, 1978.

Pflaum, S. The predictability of oral reading behaviors on comprehension in learning disabled and normal readers. Journal of reading behavior, 1980, XII, 231-236.

Rankin, E.F. & Culhane, J.W. Comparable cloze and multiple-choice comprehension test scores. Journal of reading, 1979, 13, 193-198.

Ransom, P.E. Determining reading levels of elementary school children by cloze testing. In J.A. Figurel (Ed.), Forging ahead in reading. Newark, Delaware: International Reading Association, 1968.

Richeck, M.A. Learning disabilities and remedial reading. In S. Pflaum-Conner (Ed.), Aspects of reading education. Berkeley, California: McCutchen Publishing Corp., 1978.

Rigg, P., & Taylor, E. Measures of comprehension: Independence and implications. In J.D. Harste and R.F. Carey (Eds.), Monograph in language and reading studies: new perspectives on comprehension. Bloomington: Indiana University, 1979.

Roger, J.M., Hastings, C.N., & Hook, C. A sentence verification technique for measuring reading comprehension. (Technical Report No. 137). Urbana-Champaign: Center for the Study of Reading, University of Illinois, August, 1979.

Rosenshine, B.V. Skill hierarchies in reading comprehension. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, New Jersey: Lawrence Erlbaum, 1980.

Roser, N.L. The history of reading comprehension: teaching and testing. Paper presented at the Annual Conference of the International Reading Association, Preconvention Institute #10, Atlanta, 1979.

Royer, J.M. & Cunningham, D.J. On the theory and measurement of reading comprehension. (Technical Report No. 91). Urbana-Champaign: Center for the Study of Reading, University of Illinois, 1978.

Rumelhart, D.E. Toward an interactive model of reading. (Technical Report No. 56). San Diego: Center for Human Information Processing, University of California, 1976.

Russell, S.N. How effective is the cloze technique in the measurement of reading comprehension? Viewpoints in teaching and learning, 1978, 54, 90-96.

Sadoski, M. A unitive psycholinguistic comprehension measure: the comprehension process score. In Meagher, J. and Page, W. (Eds.), Language centered reading instruction. Storrs, Connecticut: The University of Connecticut, 1980.

Sadoski, M., Page, W.D., & Carey, R. The comprehension process score: empirical testing of a new miscue scoring technique. Unpublished manuscript, undated. (Referred to in personal correspondence from M. Sadoski to L.Z. Lang, dated 3/5/81.)

Sadoski, M. (April 7, 1981). [Personal correspondence of M. Sadoski to L.Z. Lang].

Samuels, S.J. Hierarchical subskills in the reading acquisition process. In J.T. Guthrie (Ed.), Aspects of reading acquisition. Baltimore: Johns Hopkins University Press, 1976.

Sedlak, R.A., Steppe-Jones, C., & Sedlak, D. Informal assessment: concepts and practices. In T.L. Miller and E.E. Davis (Eds.), The mildly handicapped student. New York: Grune & Stratton, 1982.

Seidenberg, P.L. Implications of schemata theory for learning disabled readers. Journal of learning disabilities, 1982, 15, 352-354.

Simons, H. Reading comprehension: The need for a new perspective. Reading research quarterly, 1971, 6, 338-363.

Singer, H. & Ruddell, R.B. (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Singer, H. Active comprehension: from answering to asking questions. The reading teacher, 1978, 31, 901-908.

Smith, C.T. Evaluating answers to comprehension questions. The reading teacher, 1978, 31, 896-900.

Smith, F. Understanding reading. New York: Holt, Rinehart, & Winston, 1971.

Smith, N.B. The many faces of reading comprehension. The reading teacher, 1969, 22, 249-259, 291.

Smith, S.L. Retellings as measures of comprehension. In J.D. Harste and R.F. Carey (Eds.), Monograph in language and reading studies: new perspectives on comprehension. Bloomington: Indiana University, 1979.

Spearitt, D. Identification of subskills of reading comprehension by maximum likelihood factor analysis. Reading research quarterly, 1972, 8, 92-111.

Spiro, R.J. Constructive processes in prose comprehension and recall. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Spiro, R.J., Bruce, B.C., & Brewer, W.F. (Eds.), Theoretical issues in reading comprehension. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Strenecky, B.J., Edge, D., & Strenecky, M.K. Psychology and reading comprehension. Viewpoints in teaching and learning, 1978, 54, 43-51.

Tatham, S.M. Reading comprehension of materials written with select oral language patterns: a study at grades two and four. Reading research quarterly, 1970, 3, 402-426.

Taylor, B. Research of the reading comprehension process. Paper prepared for Seminar in Reading Research, Virginia Polytechnic Institute and State University, Blacksburg, 1977.

Taylor, B. A comparison of good and poor reader's ability to read for meaning. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, 1978.

Taylor, W.L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of applied psychology, 1947, 41, 19-26.

Taylor, W.L. Application of 'cloze' and entropy measures to the study of conceptual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.

Thorndike, E.L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of educational research, 1974, 9, 137-147.

Tierney, R.J., Bridge, C., & Cera, M.J. The discourse processing operations of children. Reading research quarterly, 1978-79, XIV, 539-573.

Torgesen, J. Memorization processes in reading disabled children. Journal of educational psychology, 1977, 69, 571-578.

Torgesen, J. Conceptual and educational implications of the use of efficient task strategies by learning disabled children. Journal of learning disabilities, 1980, 13, 364-371.

Torgusen, J.K., & Dice, C. Characteristics of research on learning disabilities. Journal of learning disabilities, 1980, 13, 531-535.

Tuiniman, J.J. Determining the passage dependency of comprehension questions in five major tests. Reading research quarterly, 1973-74, 9, 206-223.

Van Etten, G. A look at reading comprehension. Journal of learning disabilities, 1978, 11, 42-51.

Virginia Department of Education. Standards of learning objectives for Virginia public schools, Richmond, Virginia, 1981.

Virginia Department of Education, Division of Special Education, Administration and Finance. Report of handicapped children counted in local education agencies as of December 1, 1982. Richmond, Virginia, 1982.

Virginia Department of Education, Testing Service, Division of Research, Evaluation and Testing. Virginia basic learning skills testing program, Richmond, Virginia, 1978.

Vogel, S.A. Morpohological ability in normal and dyslexic children. Journal of learning disabilities, 10, 1977, 41-48.

Webber, B.L. Syntax beyond the sentence; anaphora. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Wechsler, D. Wechsler intelligence scale for children--revised. New York: The Psychological Corporation, 1974.

White, C.V., Pascarella, E.T., & Pflaum, S.W. Effects of training on sentence construction on the comprehension of learning disabled students. Journal of educational psychology, 1981, 73, 697-704.

Wong, B.Y.L. Increasing retention of main ideas through questioning strategies. Learning disability quarterly, 1979, 2, 42-47.

Wong, B.Y.L. Activating the inactive learner: use of questions/prompts to enhance comprehension and retention of implied information in learning disabled children. Learning disability quarterly, 1980, 3, 29-37.

Woodcock, R.W. Woodcock reading mastery test. Circle Pines, Minnesota: American Guidance Service, 1973.

Woodcock, R.W. & Johnson, M.B. Woodcock-Johnson psychoeducational battery. Hingham, Massachusetts: Teaching Resources Corporation, 1977.

Yeseldyke, J.E. Current practices in making psychoeducational decisions about learning disabled students. Journal of learning disabilities, 1983, 16, 226-233.

Zigmond, N., Vallecorsa, A., and Leinhardt, G. Reading instruction for students with learning disabilities. In K.G. Butler and G.P. Wallach (Eds.), Language disorders and learning disabilities. Rockville, Maryland: Aspen Systems Corp., 1982.

APPENDIX A

Appendix A: The Four Test Passages and the Cloze Practice Passage

Passage A: "Give the Ball to Wilt"

Basketball fans had known about Wilt Chamberlain for a long time. He grew up in Philadelphia. People who lived there had heard stories about a kid in junior high who could play just like the city's best college players. That kid was Wilt Chamberlain.

When Wilt was 15 he began to grow taller very fast. In only three short months he grew four inches. That means he grew more than one inch a month. As a ninth grader, Wilt was six feet eleven inches tall and still growing.

Wilt's sudden growth caused some problems at home. His father had to raise all the lights that were hanging down from the ceilings. Since nothing could be done to raise the ceilings themselves, Wilt had to walk around the house slightly bent over. He also had to sleep rolled up in a figure "S".

On the night of March 2, 1962, two pro basketball teams were warming up to play another game. But on this night the fans would be in for a surprise. They would soon see one of the best sports performances by a basketball player that anyone had ever seen before.

Thousands of fans came to watch the game. Inside the gym, they spotted one player right away. He was seven feet

two inches tall and his name was Wilt Chamberlain. Wilt made all the other pro players look like small boys.

Wilt was in this third year with the Philadelphia team. People knew he was a good player. In his first year playing pro ball, he smashed all scoring and rebounding records. He averaged an amazing 37.6 points per game. The next year, he raised this to 38.7 points per game. By his third year, he was making close to 50 points a game! Never in sports history had any other player come close to getting so many points in every game.

Wilt Chamberlain scored a record 100 points during the game that was played that night in 1962. This made him the highest scoring player in basketball history. Fans began going to games just to see Wilt play. Usually he didn't disappoint them.

Passage B: "Harriet Tubman"

Harriet Tubman was born a slave. She didn't get a chance to go to school. As a child, Harriet had to work very hard in the fields all day. That way, her owner could make a lot of money when he sold his crops. Harriet didn't think that she was being treated fairly.

When Harriet grew up, she escaped. She ran away to the northern United States. There, and in Canada, black people could be free. Harriet liked to be free. She felt sorry for all of the black people who were still slaves.

Harriet returned to the South to rescue other slaves. She helped them to escape from their owners. She made sure that they made it north, to freedom.

Harriet was in great danger because of a law that had just been passed. The law said it was a crime to help runaway slaves. She found out that slave-catchers said they would pay $40,000 to anyone who could catch Harriet Tubman.

Harriet was clever. She outsmarted the slave-catchers. She found out about a group that would help her rescue slaves. The group worked in secret. It was called the Underground Railroad. The Underground Railroad group hid Harriet and the other runaway slaves in their houses. That made it harder for slave-catchers to find the escaped slaves.

Harriet had many exciting adventures while helping slaves escape. In all, she made thirteen trips back to the South. She guided about three hundred slaves to freedom.

Then, war broke out in the United States. It was called the Civil War. The northern states fought the southern states. Harriet was on the side of the northern states because they believed that slaves should be free. Harriet worked hard for the northern army. She worked as a nurse and as a scout for raiding parties. She also spied behind enemy lines. After four years, the Civil War was over. The northern states won. All the slaves were freed.

After Harriet Tubman died, a monument was built to honor this brave woman who helped so many people win freedom.

Passage C: "First in the Sky"

Charlene awoke at five o'clock in the morning. She got dressed, had breakfast and walked to her car. It was just starting to rain as Charlene left her home for another day's work. As she drove on the highway, she thought how important it was to be careful on the slick pavement. At 6:30 a.m., Charlene arrived at work. She parked her car. Then she walked to the huge building at the airport. Charlene is an airline pilot.

There are not very many women airline pilots flying today. Charlene is one of only 110 women among the 45,000 transport pilots in America. She knows it is never easy being one of the first people to do a new job. But to Charlene it is worth the trouble. Charlene must meet all the requirements set for men who want to be airline pilots. She must have an airline pilot's license, instrument rating and a radio permit. The airlines require her to have over one thousand hours of flight time too. The airlines are very careful in selecting pilots.

Today Charlene is flying from Atlanta, Georgia, to Detroit, Michigan. She has spent many hours planning the flight with her crew. The weather tower reports that the rain will continue in Atlanta and will turn to snow near

Detroit. Charlene now checks the air traffic, fuel and the jetliner's condition.

The 727 jetliner is ready to leave Atlanta at 8:30 a.m. in the morning. Traffic control tells Charlene that the runway is clear for takeoff. The "fasten seat belts" and "no smoking" signs light up. Now the huge jet is on the runway. The pilot and crew are listening to the radio and are very busy moving the airplane's controls. The plane has gained great speed and begins climbing up and up. At ten o'clock the airplane will be landing at the snow-covered Detroit airport.

Charlene knows that an airline pilot must be alert to many things for the safety of the passengers and crew. She has worked hard to become a pilot and wants to be the best. Maybe some day she <u>will</u> be the best!

<u>Passage D: "Fine Animal Gorilla"</u>

Kolo is seven years old and can say things to people. This is very unusual because Kolo is a seven year old gorilla, not a person. Kolo doesn't use her mouth to talk. Instead, she has learned to make words by moving her hands and fingers. Kolo has learned sign language. It is the same sign language that is used by deaf people.

Some animals are good at making words with their hands. Chimps have been doing it for a long time. Experts said that gorillas were not smart enough to learn what the chimps had learned. Others felt that gorillas could not

move their fingers well enough to make words. Some people even thought that gorillas were too powerful and dangerous to teach. But Kolo proved that gorillas could learn just like chimps. She is the first gorilla that has ever learned sign language.

At first, learning to spell words with her fingers was very slow for Kolo. She learned about 20 words during her first year and a half of life. But then Kolo started to catch on. After three years of life she could make almost 200 words in sign language. Now at seven years she knows hundreds of words.

A woman named Penny Patterson found Kolo in a zoo when she was a baby. Penny became Kolo's teacher. Kolo moved from the zoo to a new home in a trailer. This is where she learned sign language. At the trailer Kolo slept in a tire just like the other gorillas at the zoo. On weekends she visited Penny on a ranch. At Penny's ranch Kolo liked to play in the trees.

Kolo is a good student. She is shy but she is friendly too. Kolo has just learned something else. Now she can use a new typing machine to communicate. She listens to people and talks back by pushing the right key on the machine. Kolo likes herself. When somebody asked her if she was an animal or a person, Kolo answered on the typing machine, "Fine animal gorilla."

Practice Cloze Passage: "The Birthday Party"

Saturdays were always special days, but this Saturday was extra special. It was Joe's birthday, and at two o'clock he was having a party. All morning he and his parents got ready for the _____. Then, at the last minute, Joe's mother remembered that had forgotten to buy candles for the cake.

The _____ was only a few block from the house, so _____ said he would go get the candles. Joe took _____ time. After all, the sun was shining, and the _____ was warm. As he walked to the store, Joe saw Mrs. Brown _____ on her front porch. Joe always stopped to talk _____ Mrs. Brown. Today he told her it was _____ birthday.

APPENDIX B: The Script of Comments and Directions to Subjects

Appendix B: The Script of Comments and Directions to Subjects

I. Examiner's Name

My name is Mrs. Lang. I am a teacher and I work at the Education Center next to W-L and the planetarium. My main job is to help other teachers learn how to teach reading better.

II. Purpose of the Study

Right now I am studying what students understand from their reading. Today and tomorrow, you will read 4 stories for me. I will check your understanding of each one in a different way. After I'm finished working here and with students at _____ and _____, I will study the results and
School School
see if I learned anything about how students understand what they read that would help other teachers.

III. Questions

A. I would like you to read this story silently. It continues on the back.

B. When you finish reading, I'll ask you some questions about it.

C. Read the story only once. If you come to a word you don't know, figure it out the best you can and keep reading.

D. Look up at me when you're finished.

(READING)

E. Here are the questions. I'm going to take notes on your answers so I'll remember them.

IV. Retelling

A. I would like you to read this story silently. It continues on the back.

B. When you're finished reading, I want you to tell me as much of the story as you can remember.

C. Read the story only once. If you come to a word you don't know, figure it out the best you can and keep reading.

D. Look up at me when you're finished.

(READING)

E. Now, tell me as much of the story as you can remember. I will run the tape recorder because I won't be able to keep up with you if I try to write down everything you say.

F. (S has stopped retelling.) That was good. Now think for a moment. Can you remember anything else?

G. Repeat if necessary.

H. (S has indicated he remembers nothing else.) You did fine on that story. Now we'll go on to another one.

V. Cloze

A. I would like you to read this story silently. It continues on the back.

B. When you come to a blank in the story, tell me a word that you think would make sense in that space and I'll write it for you on my copy of the same story.

C. If you want to skip a blank and come back to it later, that's fine; or, you can go back and change a word you've already given me if you decide you want to.

D. Here is a short story for practice.

E. If you come to other words you don't know, figure them out the best you can and keep reading.

F. Look up at me when you've finished reading the whole story.

VI. Miscue

A. I would like you to read this story out loud. It continues on the back.

B. I can learn a lot about how you understand the story by listening to you read, so I will run the tape recorder so I can listen again later.

C. If you come to a word you don't know, figure it out the best you can and keep reading.

VII. Closure

Thank you very much for helping me to study how students understand stories. In June, I will get everyone who has worked with me back together again to give a silent reading test where you will answer questions on paper. I'll look forward to seeing you then.

APPENDIX C: The Questioning, Retelling, Cloze and Miscue Assessment Data Record Sheets

Student No._____ Order No.____
School No. _____ Passage Pattern No._____
Date _____

Question Probe	Questions for "GIVE THE BALL TO WILT" (Passage A)	Acceptable Answers	Score ✓correct ✗incorrect
IM	1. Why was Wilt so famous after only three short years of playing pro basketball?	he set scoring and rebounding records; no other player had come close to getting so many points	
IM	2. How did Wilt compare to the other players on his pro team?	he scored better than they did; he was taller than the rest of the players	
EX	3. What happened during that special game in 1962 that made people call Wilt the highest scoring player in basketball history?	he scored 100 points in a single game	
EX	4. How many fans came to watch that famous game?	thousands of fans	
EX	5. In what city did Wilt grow up?	Philadelphia	
EX	6. At what age did Wilt begin to grow taller very fast?	at 15; in Junior high school	
IM	7. How do you know from story that Wilt's father cared about him?	tried to make Wilt more comfortable at home; raised the lights that were hanging down from the ceiling	
IM	8. Why did Wilt have to curl up in the shape of of the letter "S" when he slept?	Wilt was too long to stretch out full length on his bed	

Student No._____ Order No.____
School No. _____ Passage Pattern No._____

Date _____

Question Probe	Questions for "GIVE THE BALL TO WILT" (Passage A continued)	Acceptable Answers	Score ✓ correct ✗ incorrect
SCRIP	9. Why do you think the person who wrote this story decided to call it "Give the Ball to Wilt"?	the title had something to do with the fact that giving the ball to Wilt would undoubtedly result in a score	
SCRIP	10. Wilt was 6'11" tall by the time he was a ninth grader. How do you think he felt about being so tall?	probably felt very different from the other kids; probably felt like he stuck out in a crowd	

Total No. of Questions Correct_____
No. EX Correct _____(out of 4)
No. IM Correct _____(out of 4)
No. SCRIP Correct_____(out of 2)

Passage Processing Time (closest approximation in minutes) _____

Student No._____ Order No.____
School No. _____ Passage Pattern No._____

Date _____

Question Probe	Questions for "HARRIET TUBMAN" (Passage B)	Acceptable Answers	Score ✓correct ✗incorrect
EX	1. What made Harriet's childhood life so difficult and unhappy?	she was a slave; didn't get a chance to go to school; had to work hard in the fields every day	
EX	2. When Harriet escaped from slavery, why did she run away to the northern states?	black people could be free there and Harriet wanted to be free	
EX	3. Why did Harriet decide to return to the South after she ran away and became a free person?	she felt sorry for all black people who were still slaves in the South; wanted to rescue them and help them become free	
IM	4. Why were the slave-catchers willing to pay so much money to anyone who could catch Harriet Tubman?	Harriet was successful rescuing slaves and they wanted to stop her because she herself had escaped; because the law said it was a crime to help runaway slaves	
IM	5. How did Harriet keep from being caught by the slave-catchers?	outsmarted them; slave-catchers couldn't find her because she hid from them	
IM	6. Why was it dangerous to be a member of the Underground Railroad group? (Asked later, but not part of the ten questions. Was the Underground Railroad a real railroad?)	these people helped runaway slaves to escape to freedom and that was against the law in the South	
EX	7. Why was Harriet on the side of the North in the Civil War?	because people in the North believed slaves should be free; that's what Harriet believed too	

Student No._____ Order No.____
School No. _____ Passage Pattern No._____
Date _____

Question Probe	Questions for "HARRIET TUBMAN" (Passage B continued)	Acceptable Answers	Score ✓ correct ✗ incorrect
IM	8. What did Harriet do to help the northern army during the war?	she was a scout for raiding parties; a nurse; a spy behind enemy lines	
SCRIP	9. How do you think Harriet felt when the northern states won the Civil War?	happy; soon it might be possible that there would be no more slaves	
SCRIP	10. What kind of an opinion do you think other slaves had of Harriet?	admired her because she she guided so many slaves to freedom; thought she was smart, energetic, talented	

Total No. of Questions Correct_____
No. EX Correct _____(out of 4)
No. IM Correct _____(out of 4)
No. SCRIP Correct_____(out of 2)

Passage Processing Time (closest approximation in minutes) _____

Student No._____ Order No.____
School No. _____ Passage Pattern No._____
Date _____

Question Probe	Questions for "FIRST IN THE SKY" (Passage C)	Acceptable Answers	Score ✓ correct ✗ incorrect
EX	1. What was Charlene's job?	airline pilot	
EX	2. What kind of airplane did Charlene fly?	a 727; a jetliner; a jet plane; a 727 jet	
IM	3. What tells you that Charlene knew as much as a man pilot about flying an airplane?	she was hired by the airlines--that meant that she had to meet all of their requirements, same as a man	
IM	4. Why did Charlene have to be at the airport two hours before the airplane was supposed to take off?	she needed time to plan the flight with her crew, check the airplane's condition, and get weather information	
EX	5. What weatherconditions did Charlene and her crew expect when they landed in Detroit?	snow	
SCRIP	6. How could it be raining in Atlanta, Georgia and snowing in Detroit, Michigan?	temperature variations in different parts of the country caused differences in precipatation	
IM	7. Besides the crew's knowledge, what other kinds of information did Charlene rely on to fly the plane safely?	weather information; air traffic control information	
IM	8. Why were the pilot and crew so busy right before take-off?	they were listening to the radio, moving the plane's controls, turning on warning signs	

Student No._____ Order No.____
School No. _____ Passage Pattern No._____

Date _____

Question Probe	Questions for "FIRST IN THE SKY" (Passage C continued)	Acceptable Answers	Score ✓correct ✗ incorrect
EX	9. Why must an airline pilot be very alert to everything that's happening around him or her?	to maintain the safety of the passengers and the crew	
SCRIP	10. Why do the airlines have to be careful in selecting pilots?	they have to be sure that a pilot is fully qualified and experienced	

Total No. of Questions Correct _____
No. EX Correct _____(out of 4)
No. IM Correct _____(out of 4)
No. SCRIP Correct_____(out of 2)

Passage Processing Time (closest approximation in minutes) _____

Student No._____ Order No.____
School No. _____ Passage Pattern No._____
Date _____

Question Probe	Questions for "FINE ANIMAL GORILLA" (Passage D)	Acceptable Answers	Score ✓correct ✗ incorrect
EX	1. When Kolo said things to people, how did she do it?	moved hands and fingers; used sign language; used language used by deaf people	
EX	2. What kind of animal was Kolo?	gorilla	
EX	3. How many other gorillas had been taught to talk this way?	no others Kolo was the first one	
IM	4. How did Penny Patterson feel about whether or not a gorilla could learn sign language?	Penny believed gorillas could learn sign language she wasn't afraid to check out her opinion because she taught Kolo	
SCRIP	5. What do you think made Penny decide to teach Kolo after she saw her as a baby gorilla in the zoo?	Penny thought it would be good to start teaching a baby gorilla who hadn't learned anything else yet; Penny thought Kolo wasn't too dangerous to teach; Penny though Kolo was smart enough to teach	
SCRIP	6. Why do you think Kolo was moved from the zoo to a new home in a trailer?	so Kolo could concentrate better on learning the new language	
IM	7. How do you know that Kolo kept some of her gorilla habits even after she left the zoo?	no; she slept in tires and played in trees like other gorillas	
IM	8. How do you know that Penny Patterson was a good teacher?	she must have been good because Kolo made lots of progress learning sign language	

Student No._____ Order No.____
School No. _____ Passage Pattern No._____

Date _____

Question Probe	Questions for "FINE ANIMAL GORILLA" (Passage D continued)	Acceptable Answers	Score ✓correct ✗incorrect
EX	9. What do you know about Kolo that proves she wasn't a dangerous animal?	she never hurt Penny she was friendly and shy	
IM	10. How do you know that gorillas are just as smart as chimps?	because gorillas were able to learn to talk by moving their fingers and hands just like chimps were	

Total No. of Questions Correct_____
No. EX Correct _____(out of 4)
No. IM Correct _____(out of 4)
No. SCRIP Correct_____(out of 2)

Passage Processing Time (closest approximation in minutes) _____

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Give the Ball to Wilt" (Passage A)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
__ Basketball fans had known about Wilt Chamberlain for a long time. (Implied: He had played since he was a boy.)	__He (Wilt) grew up in Philadelphia.
__Wilt's sudden growth caused some problems at home.	__People who lived there had heard stories about a kid in junior high who could play just like the city's best college players.
__People knew he (Wilt) was a good player.	__That kid was Wilt Chamberlain.
__Wilt made all other pro players look like small boys.	__When Wilt was 15 he began to grow taller very fast.
__He was seven feet two inches tall and his name was Wilt Chamberlain.	__In only three short months he grew four inches.
__Wilt Chamberlain scored a record 100 points during the game that was played that night in 1962.	__That means he grew more than one inch a month.
__This made him the highest scoring player in basketball history.	__As a ninth grader, Wilt was six feet eleven inches tall and still growing.
	__His father had to raise all the lights that were hanging down from the ceilings.
	__Since nothing could be done to raise the ceilings themselves, Wilt had to walk around the house slightly bent over.
	__He also had to sleep rolled up in a figure "S".
	__On the night of March 2, 1962, two pro basketball teams were warming up to play another game.

Student No. _____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Give the Ball to Wilt" (Passage A continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__But on this night the fans would be in for a surprise.
	__They would soon see one of the best sports performances by a basketball player that anyone had ever seen before.
	__Thousands of fans came to watch the game.
	__Inside the gym, they spotted one player right away.
	__Wilt was in his third year with the Philadelphia team.
	__In his first year playing pro ball, he smashed all scoring and rebounding records.
	__He averaged an amazing 37.6 points per game.
	__The next year, he raised this to 38.7 points per game.
	__By his third year, he was making close to 50 points a game!
	__Never in sports history had any other player come so close to getting so many points in every game.
	__Fans began going to games just to see Wilt play.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Give the Ball to Wilt" (Passage A continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__Usually he didn't disappoint them.

SCORING SUMMARY

Type of Idea Unit	Number of Unit Type in Pass.	Number of Unit Type Retold	Percent of Unit Type Retold
Level 1 Idea Units (Main)	7		
Level 2,3 Idea Units (Supporting Detail)	23		
Total (Levs. 1-3)	30		

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Harriet Tubman" (Passage B)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
__Harriet Tubman was born a slave.	__When she was a child, she had to work very hard in the fields all day.
__When Harriet grew up, she escaped.	__She didn't get a chance to go to school.
__Harriet liked to be free.	__Her owner made a lot of money when he sold his crops.
__Harriet returned to the south to rescue other slaves.	__Harriet didn't think she was being treated fairly.
__Harriet was in great danger because of a law that had just been passed.	__She ran away to the northern United States.
__Harriet was clever.	__There and in Canada black people could be free.
__Harriet outsmarted the slave-catchers.	__She felt sorry for all of the black people who were still slaves.
__War broke out in the United States.	__Harriet helped them to escape from their owners.
__It was called the Civil War.	__She made sure that other slaves made it north, to freedom.
__Harriet was on the side of the north because they believed that slaves should be free.	__The law said that it was a crime to help runaway slaves.
	__She found out that the slave-catchers said they would pay $40,000 to anyone who could catch Harriet Tubman.
	__She found out about a group that would help her rescue slaves.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Harriet Tubman" (Passage B continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__The group worked in secret.
	__It was called the Underground Railroad.
	__The Underground Railroad group hid Harriet and the other runaway slaves in their houses.
	__That made it harder for slave catchers to find the escaped slaves
	__Harriet had many exciting adventures while helping slaves escape to freedom.
	__She made nineteen trips back to the south.
	__She guided about three hundred slaves to freedom.
	__The northern states fought the southern states.
	__She worked hard for the northern army.
	__She worked as a nurse.
	__She worked as a scout for raiding parties.
	__She spied behind enemy lines.
	__After four years, the Civil War was over.
	__The northern states won.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Harriet Tubman" (Passage B continued)

Level 1 Idea Units (Main)

Level 2,3 Idea Units
(Supporting Detail)

__All the slaves were freed.

__After Harriet died, a monument was built to honor her.

SCORING SUMMARY

Type of Idea Unit	Number of Unit Type in Pass.	Number of Unit Type Retold	Percent of Unit Type Retold
Level 1 Idea Units (Main)	10		
Level 2,3 Idea Units (Supporting Detail)	28		
Total (Levs. 1-3)	38		

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "First in the Sky" (Passage C)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
__Charlene is an airline pilot.	__Charlene awoke at five o'clock in the morning.
__There are not very many women airline pilots flying today.	__She got dressed, had breakfast, and walked to her car.
__Today Charlene is flying from Atlanta, Georgia to Detroit, Michigan.	__It was just starting to rain as Charlene left her home for another day's work.
__She has spent many hours planning the flight with her crew. (Implied: She prepares for take-off.)	__As she drove on the highway, she thought how important it was to be careful on the slick pavement.
__The plane has gained great speed and begins climbing up and up. (Implied: The plane takes off.)	__At 6:30 a.m., Charlene arrived at work.
__She has worked hard to become a pilot.	__She parked her car.
__She wants to be the best pilot.	__She walked to the huge building at the airport.
	__Charlene is only one of 110 women among 45,000 transport pilots in America.
	__She knows it is never easy being one of the first people to do a new job.
	__But to Charlene it is worth the effort.
	__Charlene must meet all the requirements set for men who want to be airline pilots.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "First in the Sky" (Passage C continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__She must have an airline pilot's license, an instrument rating, and a radio permit.
	__The airlines require her to have over one thousand hours of flight time too.
	__The airlines are very careful in selecting pilots.
	__The weather tower reports that the rain will continue in Atlanta and will turn to snow near Detroit.
	__Charlene now checks the air traffic, fuel, and the jet-liner's condition.
	__The 727 jetliner is ready to leave Atlanta at 8:30 a.m. in the morning.
	__Traffic control tells Charlene that the runway is clear for takeoff.
	__The "fasten seat belts" and "no smoking" signs light up.
	__The huge jet is on the runway.
	__The pilot and crew are listening to the radio and are very busy moving the airplane's controls.
	__At ten o'clock the airplane will be landing at snow-covered Detroit airport.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "First in the Sky" (Passage C continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__Charlene knows that an airline pilot must be alert to many things for the safety of the passengers and crew.
	__Maybe some day she <u>will</u> be the best.

SCORING SUMMARY

Type of Idea Unit	Number of Unit Type in Pass.	Number of Unit Type Retold	Percent of Unit Type Retold
Level 1 Idea Units (Main)	7		
Level 2,3 Idea Units (Supporting Detail)	24		
Total (Levs. 1-3)	31		

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Fine Animal Gorilla" (Passage D)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
__Kolo has learned to make words by moving her hands and fingers.	__This is very unusual because Kolo is a seven year old gorilla, not a person.
__Kolo has learned sign language.	__It is the same sign language that is used by deaf people.
__Some animals are good at making words with their hands.	__Kolo doesn't use her mouth to talk.
__Experts said that gorillas were not smart enough to learn what the chimps had learned.	__Kolo is seven years old.
__But Kolo proved that gorillas could learn just like chimps.	__Kolo can say things to people.
__A woman named Penny Patterson found Kolo in a zoo when she was a baby.	__Chimps have been doing it for a long time.
__She is the first gorilla that has ever learned sign language.	__Others felt that gorillas could not move their fingers well enough to make words.
	__ Some people even thought that gorillas were too powerful and dangerous to teach.
	__At first, learning to spell words with her fingers was very slow for Kolo.
	__But then Kolo started to catch on.
	__She learned about 20 words during her first year and a half of life.

Student No. _____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Fine Animal Gorilla" (Passage D continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__After three years of life she could make almost 200 words in sign language.
	__Now at seven years she knows hundreds of words.
	__Penny became Kolo's teacher.
	__Kolo moved from the zoo to a new home in a trailer.
	__This is where she learned sign language.
	__At the trailer she slept in a tire just like the other gorillas at the zoo.
	__On weekends she visited Penny on a ranch.
	__At Penny's ranch, Kolo liked to play in the trees.
	__Kolo is a good student.
	__Kolo has just learned something else.
	__Now she can use a typing machine.
	__She listens to people and talks back by pushing the right key on the machine.
	__She is shy.
	__She is friendly too.

Student No._____
School No. _____
Date _____

Passage Processing Time _____
(closest approximation in minutes)

Recall for "Fine Animal Gorilla" (Passage D continued)

Level 1 Idea Units (Main)	Level 2,3 Idea Units (Supporting Detail)
	__Kolo likes herself.
	__When somebody asked her if she was an animal or a person, Kolo answered on the typing machine, "Fine animal gorilla."

SCORING SUMMARY

Type of Idea Unit	Number of Unit Type in Pass.	Number of Unit Type Retold	Percent of Unit Type Retold
Level 1 Idea Units (Main)	7		
Level 2,3 Idea Units (Supporting Detail)	27		
Total (Levs. 1-3)	34		

Cloze Passage for "Give the Ball to Wilt" (Passage A)

Basketball fans had known about Wilt Chamberlain for a long time. He grew up in Philadelphia. People who lived there _____ heard stories about a kid in junior high who _____ play just like the city's best college players. That _____ was Wilt Chamberlain.

When Wilt was 15 he began to _____ taller very fast. In only three short months he _____ four inches. That means he grew more than one _____ a month. As a ninth grader, Wilt was six _____ eleven inches tall and still growing.

Wilt's sudden growth _____ some problems at home. His father had to raise _____ the lights that were hanging down from the ceilings. _____ nothing could be done to raise the ceiling themselves, _____ had to walk around the house slightly bent over. _____ also had to sleep rolled up in a figure "S".

_____ the night of March 2, 1962, two pro basketball teams were warming up to play another game. But on this night _____ fans would be in for a surprise. They would _____ see one of the best sports performances by a _____ that anyone had ever seen before.

Thousands of _____ came to watch the game. Inside the gym, they _____ one player right away. He was seven two feet _____ tall and his name was Wilt Chamberlain.

Wilt made _____ the other pro players look like small boys.

Wilt _____ in his third year with the Philadelphia team. People _____ he was a good player. In his first year _____ pro ball, he smashed all scoring and rebounding records. _____ averaged an amazing 37.6 points per game. The next _____, he raised this to 38.7 points per game. By his third year, he was making close to 50 points a game! Never in sports history had any other player come close to getting so many points in every game.

Wilt Chamberlain scored a record 100 points during the game that was played that night in 1962. This made him the highest scoring player in basketball history. Fans began going to games just to see Wilt play. Usually he didn't disappoint them.

Cloze Passage for "Harriet Tubman" (Passage B)

Harriet Tubman was born a slave. She didn't get a chance to go to school. _____ a child, Harriet had to work very hard in _____ fields all day. That way, her owner could make _____ lot of money when he sold his crops. Harriet _____ think that she was being treated fairly.

When Harriet _____ up, she escaped. She ran away to the northern _____ States. There, and in Canada, black people could be _____. Harriet liked to be free. She felt sorry for _____ of the black people who were still slaves.

Harriet _____ to the South to rescue other slaves. She helped _____ to escape from their owners. She made sure that _____ made it north, to freedom.

Harriet was in great _____ because of a law that had just been passed. _____ law said it was a crime to help runaway _____. She found out that slave-catchers said they would _____ $40,000 to anyone who could catch Harriet Tubman.

Harriet was _____. She outsmarted the slave-catchers. She found out about _____ group that would help her rescue slaves. The _____ Railroad group hid Harriet and the other runaway slaves _____ their

houses. That made it harder for slave-catchers _____ find the escaped slaves.

Harriet had many exciting adventures _____ helping slaves escape. In all, she made nineteen trips _____ to the South. She guided about three hundred slaves _____ freedom.

Then, war broke out in the United States. was called the Civil War. The northern states fought _____ southern states. Harriet was on the side of the northern states because they believed that the slaves should be free. Harriet worked hard for the northern army. She worked as a nurse and as a scout for raiding parties. She also spied behind enemy lines. After four years, the Civil War was over. The northern states won. All the slaves were freed.

After Harriet Tubman died, a monument was built to honor this brave woman who helped so many people win freedom.

Cloze Passage for "First in the Sky" (Passage C)

Charlene awoke at five o'clock in the morning. She got dressed, had breakfast and walked to her _____. It was just starting to rain as Charlene left _____ home for another day's work. As she drove on _____ highway, she thought how important it was to be _____ on the slick pavement. At 6:30 a.m., Charlene arrived at work. _____ parked her car. Then she walked to the huge _____ at the airport. Charlene is an airline pilot.

There _____ not very many women airline pilots flying today. Charlene _____ one of only 110 women among the 45,000 transport pilots in _____. She knows it is never easy being one of _____ people to do a new job. But to _____ it is worth the trouble. Charlene must meet all _____ requirements set for men who want to be airline _____. She must have an airline pilot's license, instrument rating _____ a radio permit. The airlines require her to have _____ one thousand hours of flight time too. The airlines _____ very careful in selecting pilots.

Today Charlene is flying _____ Atlanta, Georgia, to Detroit, Michigan. She has spent many _____ planning the flight with her crew. The weather tower _____ that the rain will continue in Atlanta and will _____ to snow near Detroit. Charlene now checks the air _____, fuel and the jetliner's condition.

The 727 jetliner is ready _____ leave Atlanta and 8:30 a.m. in the morning. Traffic control tells _____ that the runway is clear for takeoff. The "fasten _____ belts" and "no smoking" signs light up. Now the _____ jet is on the runway. The pilot and crew are listening to the radio and are very busy moving the airplane's controls. The plane has gained great speed and begins climbing up and up. At ten o'clock the airplane will be landing at the snow-covered Detroit airport.

Charlene knows that an airline pilot must be alert to many things for the safety of the passengers and crew. She has worked hard to become a pilot and wants to be the best. Maybe some day she <u>will</u> be the best!

Cloze Passage for "Fine Animal Gorilla" (Passage D)

Kolo is seven years old and can say things to people. This is very unusual because Kolo is a seven _____ old gorilla, not a person. Kolo doesn't use her _____ to talk. Instead, she has learned to make words _____ moving her hands and fingers. Kolo has learned sign _____. It is the same sign language that is used _____ deaf people.

Some animals are good at making words _____ their hands. Chimps have been doing it for a _____ time. Experts said that gorillas were not smart enough _____ learn what the chimps had learned. Others felt that _____ could not move their fingers well enough to make _____. Some people even thought that gorillas were too powerful _____ dangerous to teach. But Kolo proved that gorillas could _____ just like chimps. She is the first gorilla that _____ ever learned sign language.

At first, learning to spell _____ with her fingers was very slow for Kolo. She _____ about 20 words during her first year and a half _____ life. But then Kolo started to catch on. After _____ years of life she could make almost 200 words in _____ language. Now at seven years she knows hundreds of _____.

A woman named Penny Patterson found Kolo in a _____ when she was a baby. Penny became Kolo's teacher. _____ moved from the zoo to a new home

in _____ trailer. This is where she learned sign language. At _____ trailer Kolo slept in a tire just like the _____ gorillas at the zoo. On weekends she visited Penny _____ a ranch. At Penny's ranch Kolo liked to play in the trees.

Kolo is a good student. She is shy but she is friendly too. Kolo has just learned something else. Now she can use a new typing machine to communicate. She listens to people and talks back by pushing the right key on the machine. Kolo likes herself. When somebody asked her if she was an animal or a person, Kolo answered on the typing machine, "Fine animal gorilla."

Student No. _____
School No. _____ Miscue Score Sheet--Passage: _____

Miscue-Text	SC*	Graphophonemic	Semantic	Syntactic

Total raw score _____ _____ _____ _____

Maximum possible _____ _____ _____ _____

Percentage
(Total Maximum) _____ _____ _____ _____

Percentage including self-corrected _____

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2

THE RELATIONSHIPS AMONG ALTERNATIVE MEASURES OF READING COMPREHENSION IN LEARNING DISABLED STUDENTS

by

Lynn Z. Lang

(ABSTRACT)

Literature on reading comprehension assessment advocates a multiple-measures approach over a single-test methodlogy in view of the cognitive and linguistic complexity of the behavior being evaluated. Relationships among multiple measures have been observed, but no study has examined the nature of relationships among categories of comprehension assessment (product-process) or has conducted an explicit test of possible patterns of association, expressed as three models of comprehension assessment, in a sample of learning disabled subjects.

Therefore, the passage comprehension of 48 seventh and eighth grade males was assessed by questioning, retelling, cloze and miscue assessments followed by a correlational analysis of scores. Regression analysis examined which of the four assessments best predicted subjects' scores on a standardized test.

The data did not confirm any of the three hypothesized models of comprehension assessment but subjects' comprehension appeared to be partly a function of how it was measured. Performance on the miscue assessment best predicted standardized test performance.