DOCUMENT RESUME

ED 218 950 FL 013 029

AUTHOR Bauman, James
TITLE Linguistic Structure and the Validity of Reading Comprehension Tests. Final Report.
INSTITUTION Center for Applied Linguistics, Washington, D.C.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE May 82
GRANT NIE-G-80-0149
NOTE 121p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Reading Comprehension; Reading Research; *Reading Tests; *Standardized Tests; *Structural Linguistics; Testing Problems; *Test Validity; Test Wiseness

ABSTRACT

This study considers whether or not the various sentence structures in standardized reading tests cohere to facilitate the extraction of information. In the context of the theoretical framework of the study, a schema theory of reading, a diffuse discourse structure may legitimately activate more than one way of conceptualizing an answer to a test question. The methods used to explore this issue were the collection of a set of clinical interviews with children in grades 3, 5, and 8 in the Montgomery County (Maryland) Public Schools who had taken a standardized achievement test; and the administration of an experimental test. The interview data indicated a strong probability that test performance could be influenced by the structural organization of passage information. Analysis indicated that low-achieving students lacked metacomprehension and test-taking skills, not comprehension ability. An experimental test was designed to reduce ambiguities, inconsistencies, and inaccuracies in the passages and questions in order to assess more accurately students' comprehension ability. An important result was that the main difference between good and poor readers involves a reluctance or inability on the part of poor readers to process information hierarchically. (AMH)

Linguistic Structure and the Validity of Reading Comprehension Tests

Grant No. NIE-G-80 0149

Final Report

to.

National Institute of Education

James Bauman, Principal Investigator

Center for Applied Linguistics, Washington, D.C.

May, 1982

Preface

Notwithstanding the title of this report and its failure to mention reading instruction, the conception and design of this study originated in a practical educational problem. How closely did the demands of a particular reading comprehension test match up with the demands and objectives of a particular reading curriculum? The curriculum in question was the Instructional System in Reading/Language Arts of the Montgomery County (Maryland) Public Schools which over the years of its development has stressed the importance of detailing objectives based on the discourse properties of reading passages. This is an innovative approach to reading curriculum development and one which it was not certain would align with the objectives of standardized reading comprehension tests. The test objectives, while admittedly not ignoring discourse concerns, are broadly stated in terms of modes of processing which presumably crosscut the various types of discourse included on a reading test.

It became clear quite early into the investigation that the MCPS objectives were not being measured on the standardized test--a not surprising conclusion in itself, since the intent of the tests was at variance with the intent of the curriculum. There is, of course, no value judgment that can be made on the simple evidence of a mismatch. However, in honing the analysis, there was reason to suspect that the tests had invalidated certain questions by not attending to matters of discourse organization. In other words, there were indications that a child taking the test might legitimately see more than one right answer or might legitimately eliminate the correct choice.

What constituted a 'legitimate' problem, however, was at that stage of the inquiry a matter of some uncertainty because the theory capable of informing a sure judgment was itself imprecise. It was in the interests of contributing to the clarification of this theory that the present study was proposed to the National Institute of Education. The intention was to expand the scope and purpose of the MCPS investigation, first to subject several reading tests to a fine grained linguistic analysis, focussing especially on discourse properties, to question test takers about their behavior and impressions on a test they were required to take, and, last, to formulate and test a set of hypotheses about what constituted imprecision and difficulty in the tests. The tests selected

were the California Achievement Test (CAT) (1978 edition, Forms C and D, Monterey, CA: CTB:McGraw-Hill), the Metropolitan Achievement Test (MAT) (1978 edition, Form JS, New York: The Psychological Corporation), and the Iowa Test of Basic Skills (ITBS) (1979 edition, Forms 7 and 8, Iowa City: Houghton Mifflin).

The study was carried out where it had been initiated, in the Montgomery County Public Schools, and I would like to express my gratitude to the district personnel and the principals of the schools cooperating in the study for their interest and help. I am especially thankful to Ted Schuder and his staff in Instructional Development for paving the way, lending an ear, and benfitting me with their breadth of experience in reading and testing and to Susan Gross in Accountability for her help with the data analysis.

The project was implemented with the invaluable assistance of Joan Koppelman, who consistently and with dedication performed beyond my expectations. Besides skillfully managing the considerable number of clerical chores of the project, she acted, out of her background as a teacher, as a valuable brake on my speculation. She came ultimately to have an important influence on many of the ideas which became embodied in this report. In Joan's absence during the preparation of this report, Donna Christian and Ruby Berkmeier ably and with considerable patience took on the burden of formatting, proofing, and typing, for which I am deeply appreciative and in their debt.

I owe, of course, other debts of a less tangible nature to other investigators in reading and testing. I have acknowledged some of these in the context of the background discussion, but many others should have been included. It was in the interests of keeping the discussion as focussed as possible on the documentation of the study that greater consideration was not given to more of the relevant literature. Beyond this consideration, however, a good deal of pertinent research -- in particular, a study being conducted by Paul Kay and Charles Fillmore (see Fillmore 1982)--is still in progress, and a comparison of findings would have been premature. In later publicaitons and as appropriate this work will be properly referenced and discussed. Hopefully, by that time, this report will have accomplished its interim task of encouraging the kind of critical response needed to sharpen its presentation and expand the scope of its educational implications.

Table of Contents

INTRODUCTION

The ideal test knows in detail what it is doing. It knows that its questions fall within the range of experience of the people who will take it, it understands the perceptual and cognitive demands of its questions, it encourages only legitimate applications of its results, and it looks like what it claims to be. Collectively, these considerations constitute matters of test validity, reassurances that the test does in fact do what it claims to do. The first consideration, for example, concerns the matter of content validity. The test questions require answers which the test takers can with some probability be expected to know. The second consideration is more abstract and harder to discern. It assumes that the test has been constructed in conformity with generally held beliefs about how similar information would be processed in a non-test situation. Furthermore, it assumes that the test has been fairly constructed, at a minimum presenting no ambiguous or misleading alternatives. Considerations of this kind are encompassed in the term construct validity. The third consideration requires that the test make claims about its usefulness that are consonant with its objectively determined value. This requirement in effect restricts the scope of predictions that can be made about the test taker's behavior on the basis of test results. This consideration is generally referred to as predictive validity. The last consideration constitutes the weakest demonstration of validity. It requires only that the test be formatted so that test takers recognize it as a test of what it claims to be, a matter of what is generally called face validity.

A carefully constructed test instrument will take pains to demonstrate to users that it has satisfactorily addressed each of the validity considerations. Part of this demonstration typically entails the presentation of evidence that the test on repeated applications achieves statistically comparable results, an indication of its reliability as a test instrument. A reliability check, however, does not of itself fully answer questions of validity. Most importantly, we would still not know that the test actually measures what it claims to. It might be reliably measuring some aptitude or behavior which the

test makers had not anticipated. In other words, although a reliable test has validity for some purpose, it may not be the purpose that the test makers have in mind.

Checks on test validity typically involve a correlational analysis. For instance, if the concern is the content validity of a reading test, the test developer might want to show that the instrument resembles in important respects the curriculum used to teach children to read. If the concern is the test's predictive validity, the test makers would look for strong correlations between test performance and future academic success such as grade advancement or school leaving. If the concern is face validity, the test maker might want to show that test results correlate highly with results on other tests intended for the same purpose.

Checks for construct validity are probably more difficult to make than other validity checks, although in principle the methods are similar. The test maker still wants to find a correlation, but searches for one between the skills used in solving reading test problems and the skills used in other reading situations. What test makers usually do is engineer the test in such a way that individual questions address particular test objectives, these being a set of comprehension or problem solving skills. The Iowa Test of Basic Skills (ITBS), for instance, organizes its reading comprehension subtest around sixteen skills, each falling within one of three more inclusive categories. For example, individual questions are determined to reveal information on the test taker's ability to understand factual details relating to classification, a literal meaning skill; to draw conclusions from information and relationships, an interpretive meaning skill; or to recognize the main idea or topic of a paragraph or selection, an evaluative meaning skill. Thirteen other skills are tested in addition. Even from this sketchy description, it is apparent that the delineation of skills objectives and the demonstration of construct validity are intended to be useful for classroom instructors and curriculum developers. If a test lays claims to being a instrument able to help educators in making instructional decisions, the worth of the claim lies substantially in the demonstration of the test's construct validity.

The other demonstrations of validity are perhaps of more importance to administrators. Issues revolving around content validity, for instance, ultimately affect decisions about whether to favor one test over another. Obviously if a test is constructed along the same lines as the curriculum, it will be to the school's advantage to select it over another test constructed along different guidelines. Then too, if the school is primarily interested in using the test results to aid in student placement decisions or to compare student performance, more decision weight might be given to the demonstration of predictive validity.

The complexity of validity considerations leads us naturally enough to question how much any one test can hope to accomplish and how many interests it can hope to serve. Many researchers in the testing field have pointed out the trade-off which must be made between a test capable of discriminating individuals and a test capable of diagnosing individual strengths and weaknesses. The problem centers around the point at which a particular test question becomes difficult to a majority of test takers. For purposes of comparing and ranking a set of test takers, it is best to have the average difficulty of test items center around 50%. At this rate half the test takers would answer the item correctly and half incorrectly, and the performance of all test takers would approximate a normal distribution. In this way, the test can maximally discriminate among all test takers. At a higher difficulty level the test has lesser power to discriminate among poor achievers, while at a lower level it loses its ability to discriminate among high achievers. On the other hand if the purpose of the test is not to compare test takers but to evaluate the standing of each test taker against some desired standard, say satisfactory completion of 80% of the test items, then difficulty levels of individual items must be lowered so that the test as a whole approximates the standard in difficulty. The logic dictates that a test designed in such a way that half of the students correctly answer less than half the items has little power to inform evaluators about the true or absolute achievement of individuals. The logic is predicated on a culturally based assumption that a group level of achievement no higher than 50% borders on being deplorable and indicates an unacceptable level of

effort. If the test has an educational intent and group performance is as low as 50%, someone invariably has to provide an explanation of why performance is not higher.

Up until 20 years ago educational testing in this country consisted primarily of teacher-made tests, seeking to reveal children's understanding of their instruction, and standardized norm-referenced tests, seeking to inform teachers and parents about how well children were doing in comparison to one another. The teacher-made tests, being specific to the instructional goals of a single teacher, were not subjected to validity or reliability checks. It was necessary to accept their validity on face value, trusting to the integrity and credibility of the teacher. The standardized tests, with their aim set at the performance of students across the nation, obviously could not be tied closely to the instructional objectives of any one curriculum. To demonstrate their applicability to a wide ranging set of users, careful checks on reliability and validity had to be conducted. The effect of these checks was to confer on the tests the impression of objectivity, a distinction which the teacher-made tests might claim but could not prove. Over time this led educators concerned about the accountability of their actions to put greater reliance on the standardized tests and make them serve double duty as diagnostic and placement tools. From the standpoint of test theory, however, this extension of scope was questionable since by design half of the students taking the test would score below average and would look like they needed remediation. But do they in fact need remediation or is it just that the score of the average achiever is set unrealistically low?

Possibly on theoretical grounds alone, but more likely because of the potential for abuse which arises when an instrument is used for an inappropriate purpose, a reaction set in against the use of standardized norm-referenced tests for making decisions about individual students. Charges were brought that the tests were culturally biased, that they were tailored for middle class white American values and experiences and that they discriminated against members of other groups. Other charges were made that the tests were being administered to groups who were excluded from the norming sample--in a documented case

Spanish speaking students with limited English ability--and that the test results were used as justification for placing these students in classes for the mentally retarded. Overall, however, the major criticism came to be that the label a student acquired because of performance on a standardized test endured, not so much because there was truth in it but because the placement decisions made for the student perpetuated it. Students said to be below average are taught as if they were below average. Over time they come to believe the label and set their aspirations in accordance with it. The feeling is growing that too much import hangs on the student's performance on a single achievement test, especially given that factors such as test anxiety, motivation, interest, and expectation also enter into the equation which produces the final score.

With standardized tests as influential as they are in informing educators, it is critically important to view them as objectively as possible and to evaluate the strength of their claims against a reasoned and balanced assessment of their design and implementation. In the study reported here we will look at a fundamental issue on which will turn a judgment of the construct validity of standardized reading comprehension tests. Does the linguistic structure of reading tests influence the comprehensibility of test items? This question has a trivially affirmative answer under the meaning that any test passage or question must conform to an acceptable standard of grammaticality. It is not, however, this level of analysis which primarily concerns us. On the surface at least, most sentences in reading tests are well-formed and structurally intact. Less concedable is whether the various sentence structures cohere to facilitate the extraction of information. It is possible that they may be so disjointed that information integration is impeded. In the logic of the schema theory of reading adopted here (Chapter 1), a diffuse discourse structure may legitimately activate more than one way of conceptualizing an answer to a test question. If this happens then skills other than those strictly involved in reading comprehension are involved in selecting the 'correct' or expected answer.

In Chapter 2 a discussion of the methods used to explore this issue are described. They include two major procedures, the first a set of indepth, clinical interviews with children in the 3rd, 5th, and 8th grades and at two proficiency levels. The point of the interviews

is to determine how and why children taking standardized reading comprehension tests answer the way they do. The results of the interviews are discussed in Chapter 3.

The interview data, although it proved valuable in its own right, was originally intended to serve as a conduit to an experimental procedure by suggesting ways in which test passages could be rewritten to enhance their comprehensibility and improve test performance for low achieving students. The interviews did in fact indicate a strong probability that test performance could be influenced by the structural organization of passage information and the experiment was designed and executed as planned. The results of it are presented in Chapter 4.

The sum of the information presented in this report strongly suggests that greater attention be paid to the design of reading test passages and questions. Two points in particular are argued: the tests in the form they now exist are inadequate to assess the true comprehension skills of the children who take them; moreover, they fail to assess the actual proficiencies by which children achieve or fail on individual items. Each of these claims constitutes a serious doubt regarding the claimed construct validity of the reading tests. A more detailed synthesis of the argument and its ramifications is presented in Chapter 5.

CHAPTER 1

THEORETICAL FRAMEWORK

The choice of an appropriate level at which to conduct a linguistic analysis of reading tests has to be motivated by the full variety of objectives which the tests purport to treat. Since many of these objectives presuppose a data base of greater extent than a single sentence, it would be advantageous to direct attention to the linguistic structures which characterize the discourse or text as a whole. This point is hardly argued anymore. Most researchers in reading and testing, in fact, do concern themselves with discourse processing issues, although there is no large scale agreement on what constitutes a viable approach to discourse analysis or even on how to define discourse (Winograd, 1977). The differences result from researchers beginning their analyses from different points, some from the level of the sentence working up and others from the level of discourse purpose working down. Part of the difficulty stems from the wide scope of discourse, which as a field encompasses the full breadth of concern for meaning in language. Discourse analysis, consequently, has been broached by rhetoricians, linguists, philosophers, folklorists, anthropologists, psychologists, sociologists, and computer scientists, each group addressing the concerns arising from its own discipline and applying methods of analysis which are favored or even idiosyncratic to its own field. The full integration of these various approaches has not as yet been accomplished and as a result we are still operating without a comprehensive discourse theory.

It was beyond our scope to engage the full range of interdisciplinary issues involved in discourse studies, although we did consider those issues which collectively would lead us to formulate a working definition of discourse. The major desideratum was to characterize discourse as a self-contained structural entity not defined simply as an aggregate of sentences or as an exemplar of some communicative intent. The proper defining characteristics were felt to be those structure-to-function analogs from which writers select to give their accounts coherence and cohesion, to achieve their

purposes in writing, and to satisfy the needs of the reader. Implicit in this characterization are the existence of an interpersonal interaction between the reader and the writer and a constructive interaction between the reader's background information and the new information in the passage.

Writer-reader interaction

In certain respects the writer-reader interaction has characteristics in common with conversational interactions, differing mainly in that the writer and the reader are not in temporal or physical proximity at the time of the communication. The commonality of the two events hinges on the structural implications which follow out of a need for the communicants to be cooperative. Grice (1975) first described these conversational implicatures, as he calls them, as predicated on a Cooperative Principle, whereby each participant recognizes in a conversational situation "a common purpose...or...a mutually accepted direction" (45). The stated or implied agreement on the purpose of the communication makes judgments about the acceptability or appropriateness of individual remarks possible. If participants are communicating in good faith, then both assume that any remarks made are implicated in the purpose of the conversation. In other words, everything that is said is expected to have some discernible relationship to the reason the principals are communicating.

We assume that the same postulates hold for a written communication, although the interaction of the writing situation is necessarily less dynamic than its conversational counterpart. In a face to face conversation, the information exchange can be individually tailored to the expressed needs and informational backgrounds of both participants, while in a writing situation writers may not know with certainty who their readers will be. They have to anticipate the information needs and backgrounds of possibly a wide variety of readers, who each may be reading for different purposes. Conversational encounters typically include structural devices whose purpose is to assure that both participants are perceiving each other's information equivalently. If one of the participants has difficulty identifying the function of a particular utterance, the other participant can interrupt the flow

of the conversation to its determined purpose long enough to show the relevance of the remark. Keenan and Schieffelin (1976) demonstrate how this procedure works even at fundamental levels, for instance in situations where both participants must assure each other that they have the same topic referent in mind. Writers are less able to resort to such devices because they are providing a monologue account and because, not being in face to face contact with their communicants, they lack immediate feedback data from non-verbal signs (Winograd 1977). The puzzled looks, squirming, and signs of inattention, which would otherwise inform them about how well their material was being comprehended, are absent. The decisions that a writer makes must then be more considered or planned (Ochs 1979) if they are to be effective in communicating the intended message.

But in what does the planning consist? What sorts of criteria must the writer turn around mentally and transpose onto paper to make the message salient and clear to the reader? Intuitively, approaches which take no recognition of the purposes for which people engage in literacy events or which take no recognition of the structure in which information is conveyed cannot provide the answers to these questions. The answers will be found at a level of analysis where structure intersects function and where the continuity of the message can be detected. The appropriate starting point of the study was therefore determined to be neither at the level of the sentence nor at the level of the discourse as a whole, but at the intermediate level of the paragraph.

The paragraph as a unit of discourse

The status of the paragraph as an organizational entity has been argued primarily out of a rhetorical tradition. Rhetoricians, however, have almost exclusively approached the paragraph from the perspective of function and style.

Their concerns have been with optimizing information flow; describing how ideas can be presented to get a message across with no loss of accuracy. They have paid lesser attention to the linguistic mechanisms underlying the accomplishment of these goals. Consequently the terms which have been elaborated out of the rhetorical tradition--

topic, comment, reinforcement, illustration, explanation, evaluation, and so on--have primarily functional definitions. We know little from traditional rhetoric about the structural realization of a category such as the paragraph topic, other than the position it typically assumes in the paragraph. Lacking a structural focus, the rhetorical tradition has not been able to convincingly describe the linguistic characteristics of a paragraph; for instance, it is difficult to see what structural commonalities might unify the topic category.

Recently, however, rhetoricians have begun to approach the analysis of the paragraph from the reader's point of view, a perspective which promotes questions of a structural nature. Underlying this point of view is a conception of the writing event as being fulfilled in a reading event and implied is an interaction between the writer and a reader. Young, Becker, and Pike (1970), instance, describe paragraphs as structures which "arouse and fulfill anticipations" (324) in the reader's mind. Although these authors do not use the terminology of reading theory, their conception is similar in all important respects to the notions encapsulated under schema theory (see below). Writing and reading imply a sharing of information which is successfully accomplished only if the writer and the reader acknowledge each other's backgrounds, values, and social relationships. With no basis for sharing, no information transfer will occur.

From the other direction, some linguists working within the field of discourse analysis are also recognizing the integrity of the paragraph as a grammatical unit. Longacre (1979), for instance, has described several languages in which there are formal grammatical and lexical markers of paragraph boundaries. He believes that these devices manage the integral function of maintaining the unity of information relative to a thematic participant in narrative (and presumably a thematic referent in other forms of discourse). At the point where events change the identity of the main participant, one paragraph will have been concluded and another one begun. Longacre continues by positing three structural parameters--concerned with logical structure, cohesive relations, and stylistic factors--whose

various constituents interact to form a typology of paragraph genres and paragraph transitions.

Working within a recall paradigm, Chafe (1979) has also come to posit the identity of the paragraph as a structural linguistic entity. He had subjects view a short, silent film and later report back what they had seen. The reports, which were analyzed for duration of transitions between utterances, demonstrated that long pauses and hesitations were reliably associated with informational "breaks in the coherence of space, time, characters, events, and worlds" (180). Chafe refers to these breaks as paragraph boundaries presumably because at these points his subjects had concluded their information relay concerning one topic focus and switched to another.

Interaction with background information

The second premise of the study, that readers construct their comprehension of a passage using their accumulated store of knowledge, has a long history in the psychological literature. Rummelhart and Ortony (1977) take the concept back to Kant, who talked of a "productive imagination" characterized by an ability to apply its categories of meaning to the realization of knowledge. Kant referred to the rules governing this ability as schemata (singular schema) a term which has been revived periodically in the psychological literature and which currently forms the conceptual basis for a wide range of studies in reading and cognition. As Rumelhart (1980) defines it, schemata are the building blocks of cognition "the fundamental elements upon which all information processing depends" (33). In application to reading, they determine the reader's goals in reading, organize the processes of retrieving information from memory, and guide the organizational path through the material. In other, related views they are cognitive structures comprised from background information and experience against which new information is compared and assessed for meaning and relevance.

These conceptions of schemata have been called upon to explain the reading comprehension process in both its successful and unsuccessful aspects. In successful comprehension readers are said to engage a pre-existing schema and use it to process the written

material before then for whatever purpose they are reading. Evidence of successful recall is most commonly adduced through demonstrated ability to recall information read or otherwise perform in a way dictated by the written material, say in accomplishing a procedural goal. Comprehension is said to fail if the reader is incapable of recalling the information or of demonstrating the required level of proficiency. In these cases the problem is attributable to some unsuccessful attempt on the reader's part to engage a schema. In the most obvious conception of the difficulty, readers may fail to understand simply because they lack any familiarity with the topic; they have no schema to engage. In more complex conceptions (Spiro 1980), readers' failure at comprehension can be attributed to an inability to access a relevant schema, either for some reason buried in their motivation or approach to the situation or because the author does not make it sufficiently obvious what path to take to the schema.

The selection and application of a schema to a reading task implies that the reader is striving to reach some learning goal. Adapting the categories layed out by Spiro (1980), several outcomes are possible. The reader could relate the new information across a series of comparable past literacy events, resulting in an updating of the accessed schema. This might occur, for instance, when the reader learns in the established context of following a baseball team's progress that it won a game the preceding day. This is equivalent to accumulating facts on a particular topic. In a second outcome, the reader may expand a schema by incorporating new information, for example, by learning that whales and porpoises are mammals. This type of learning by itself results in additions to categorical knowledge. If, however, the extended effect of the additional information about whales is a reformulation of the reader's conception of mammals and fishes, then the net learning result is a schema alteration. Lastly, in the null case, where a schema is not available, a form of learning which Spiro calls compartmentalization may nevertheless take place. Learning of this type essentially involves the introduction of a new schema, which, being untied to any background information, probably has a rather tenuous existence in the reader's long-term memory storage. Other, higher order learning effects in which schemata may be amalgamated,

hierarchized, rationalized, evaluated, and so on are also possible results of a reading event, but it is likely that they involve complex reflective activities in addition and perhaps occur less immediately to the actual reading situation.

In its accepted sense, learning in a school or testing context is thought to be synonymous with schema alteration and the ability to perform higher order interactions of schemata. When these skills are tested for, however, it becomes apparent that students are often incapable of performing at levels comparable to the expectations of the curriculum. The usual explanation is that students in this category are developmentally below their high achieving peers, the premise being that development precedes learning. L.S. Vygotsky (1978), however, has reacted to this established notion by emphasizing the skills which children bring with them to their first classroom experience. He sees children as possessing much of the background knowledge and skills necessary to perform reading tasks within their proper situational domains. Paraphrasing his position within schema theory, he sees children as possessing many well developed and integrated schemata which they could potentially apply to their classroom and test-taking experiences. To explain the apparent developmental lags that tests suggest some children have, he advances the opinion that the instructional routine of schooling is concerned largely "with the assimilation of the fundamentals of scientific knowledge" (1978:90), and not as concerned with the steps preparatory to assimilating those fundamentals. The usual steps to discovery, he emphasizes, are observation, repeated exposure and practice, conversation with one's peers, and consultation with those knowledgeable on the topic. In a turnaround from the pedagogically entrenched position that development precedes learning, he states that "organized learning results in mental development and sets in motion a variety of developmental processes that would be impossible apart from learning" (1978:90). Cognitive development, in other words, is critically dependent on previous learning.

In the present study we hypothesize that the structure of idealized reading material is organized in such a way that readers are led naturally through the processes which result ultimately in a schema alteration. The writer in effect has made numerous decisions about what his intended reader knows and lays out his thoughts and

sentences accordingly. In the scheme of Vygotsky's learning theory, each of the author's decisions has implications for guiding the reader from one stage of learning to the next. In order to produce an alteration in a reader's schema, the author must gain the reader's attention, alert what interest is already there, provide a convincing argument, and perhaps make the point of the argument explicit. The reader, who presumably began the reading with a rudimentary schema, responds first by selecting the appropriate schema, updating it by acknowledging the author's topic as relevant, assimilating additional information to it, and then restructuring it or not as seems appropriate.

Many instructional materials and many instructional strategies fail to accomplish what they are intended to do. But since reading is an interactive process, it would be unfair to blame either the reader or the writer, without first considering the contribution that each makes to comprehension. In the next section we will lay out a scheme for considering the adequacy of the written word in conveying a message. In a later chapter we will reverse the focus and consider the reader's role in processing the message.

Most of the previous work in the linguistic analysis of discourse can be categorized into three broad and somewhat overlapping types: functional analyses which begin with a categorization of the semantic componentry of discourse (Grimes 1975, Meyer 1975); propositional analyses which start by breaking down the logical structure of the sentences composing the discourse (Frederickson 1975, Kintsch 1977); and cohesion analyses which focus on the linguistic devices linking sentences and propositions in the discourse (Halliday and Hasan 1976, Fillmore 1974). The basis for choosing one or another of these methodological perspectives lies fundamentally in the underlying belief of the particular researcher that language is inherently organized in a particular way. The propositionally-based analyses, for instance, tend to regard discourse as governed by interrelated rules, making it conformable to formal grammatical methods of analysis. Out of this tradition there have been elaborated analytical constructs such as story grammars (Stein and Glenn 1978, Rumelhart 1975);

scripts (Schank 1975); frames (Winograd 1975); and so on. The intent of all of these treatments has been to advance formal models of discourse structure. Many of the researchers in this tradition, in fact, have taken their analysis as a means toward the larger goal of designing 'intelligent' computer systems. Category-based methods are less concerned with formalisms and more concerned with describing the full extent of the informational complexity involved in discourse. The starting base for such studies may be a comparative treatment of several or many languages, the intent at the initial stages being to contrast the structural organization of information in different languages and at later stages to uncover the common basis of information exchange across languages.

It is perhaps somewhat unfair to claim that proponents of either the propositional or category-based methods are unconcerned about the distinction between good and poor discourse since the starting point for much of their research is how to distinguish random strings of sentences from true text. And yet, their methods are basically insensitive to this question. Both groups, for instance, take as their data base actually occurring discourse, making no 'prescriptive' judgments concerning its well-formedness. The underlying premise adopts the position of the producer of the discourse and states that whatever he or she regards as communicative is, in fact communicative. Given the premise that discourse interactively involves both communicants, any judgment concerning well-formedness lies with the reader/hearer as well as the writer/speaker. It is important to inquire into what the reader can and must do with the information as presented in order to reach the state of understanding the writer has intended. Questions of this sort, it would seem, are more immediately answerable by invoking methods which concentrate on the processing demands of prose. These, we feel, are more directly addressed by studies focused on the cohesion relations tying linguistic units together.

In this tradition, Halliday and Hasan's (1976) seminal work on semantic cohesion takes the position that a string of sentences can be distinguished from a true text because the text will exhibit a certain

register and a definable cohesion. The first characteristic places the discourse in an anchored setting and affective environment, the second assures that the informational relay is motivated, consistent, and targeted. Although Halliday and Hasan themselves do not directly address the question of well-formedness, they assemble and blueprint the notational machinery necessary to discuss it--anaphora, conjunction, lexical synonomy, ellipsis, and substitution.

Informational organization in discourse

It is rarely if ever denied that the organization of information in text influences comprehension. Writing which is poorly constructed taxes the processing capabilities of readers. It confuses the reader by juggling a point of information in ways which make its relative importance unclear and its position in the discourse contrary to expectation. And yet the widespread concern for improving the quality of writing, especially for children, has not yet yielded hard evidence indicating how this might be done. Most of the current claims have been made by researchers working within experimental psychology using recall paradigms. The results of this research focus have almost always supported a model of comprehension based on a hierarchical processing of discourse. Meyer (1977), for instance, has demonstrated that readers selectively and preferentially recall information which an independent discourse analysis would tag as critical to the continuity of the reading passage. In other words, readers tend to recall the 'thread of discourse' (Grimes 1975), that is, the main ideas, the topics, the key points, and whatever else they have been called, better than they do the supporting details of the passage.

These results seem intuitively satisfying, possibly because they confirm the impressions we receive out of our educational training that a construct such as main idea is a relatively more important component of text organization than is a supporting detail. The semantics of the terms themselves suggest this. And yet the conclusions reached through this line of research have not gone unchallenged. Spiro (1977) and Olson (1977), for instance, have argued that the recall task used in these experiments constitutes an unreliable measure of reading comprehension in that it fails to take into account the reader's background

or the purposes for which individual readers may undertake a reading task. They argue against the notion of hierarchical processing and for a position in which the reader has the flexibility to pick and choose from among the full range of information available within a passage. The actual choice of material deemed important would depend not so much on any 'inherent' organization of the text as it would on the information needs of the particular reader. As Olson says, a reader "comes already equipped with a very elaborate set of expectancies, a host of prior cues, and a set of values as to what is worth looking for,.." (212).

These arguments have lately been reiterated by other researchers. Trabasso (1981), for instance, suggests that sentences which perform the main idea function in discourse are in their very structure easier to commit to memory than other sentences. This comment implies that there may be specialized semantic and syntactic structures associated with identifiably different discourse roles, an idea that is itself akin to syntactic structures, such as relative clauses, direct objects, passive constructions, and so on, having their own intrinsic meaning. It further implies that there should be ways in which to assess comprehension other than to use recall procedures. Baker and Stein (1981) concur on this point, and suggest that more appropriate alternatives would be to assess comprehension by looking at how information is applied or integrated into other knowledge. Conventional reading tests do make these kinds of demands on readers, but the significance of the demands in particular instances is difficult to evaluate without a better appreciation of what processing requirements are entailed.

Baker and Stein (1981) suggest that for narrative discourse--material which essentially relates a set of temporally connected happenings or events (Labov and Waletzky 1967)--the story grammar model could provide a useful guide to assess knowledge integration. Their claim does not ignore the difficulties cited above regarding the use of formal models of discourse as psychological models of comprehension; it tempers them by suggesting they are more relevant to expository discourse--material which "is typically abstract, dealing with unfamiliar concepts and situations" (41). Their claim for narrative

discourse is that it presents a structure which is highly determined through social convention, thereby exhibiting a certain logic, which is apparent even to young children.

Categories of information in discourse

The consensus of the literature as we have interpreted it is that there does exist a structured quality to discourse which guides the reader or listener to the intended meaning of the writer or speaker.

The structure is not fitted de novo to each instance of newly created discourse; it is to some extent culturally agreed upon. In essence this means that the writer and reader share an understanding about how the information in the full message should be organized and presented. The depth of the understanding also seems to vary with the larger purpose of the discourse, the conventions being more certain for narrative than for exposition. In addition the reader must have in mind the necessary schema to understand the content of the message. At a minimum this requires the reader's understanding of the referents the author takes for granted. Finally, the proper level of analysis is the paragraph unit, rather than the sentence or the discourse as a whole. We presume that it is at this level where structure and function optimally interact.

The given in any discourse exchange is that the originator has something--some referent--in mind to say something about. This referent and the preliminary statement made about it is conventionally referred to as the topic, the topic sentence, the topic referent, or the paragraph topic. The topic has also been identified as the main idea, although main ideas are also associated with explicit or implicit statements of the originator's intent. The main idea of an Aesop's fable, for instance, might be said to be the concluding moral. In a wider sense the main idea of a discourse can be any idea which the reader so declares it to be, based on the notion that readers' needs vary and so may not be met by the same sentence for any two people. The currency of the term main idea in the testing profession seems to reflect a pedagogical reality rather than a discourse reality.

The second reality of a discourse exchange is that the topic will be elaborated in some detail. Details logically falling within the

scope of the topic are offered not necessarily just for their pure information value but also to implicitly communicate the writer's purpose. The detail, then, more than the topic determines the type of discourse; at the first cut whether it will be a narrative or expository account. At finer levels of analysis the expository category can be subdivided into categories such as explanation, process, argument, demonstration, elucidation, and so on, and the narrative category into subtypes such as dialogue, recollection, and portrayal. Certain types of detail, such as description and characterization may equally well serve either a narrative or expository purpose and muddy the border thought to distinguish the two superordinate categories.

Detail should in its logical structure, if not in its actual choice of vocabulary or grammatical structure, fall within the scope of the topic. That is, the informational content of the detail sentences should legitimately qualify as information pertinent to the topic. Since, however, the detail also veils the discourse purpose, it often happens that the originator will make this purpose explicit in a statement following on the detail. This statement will be here called an assessment. It represents rather than a necessary category of discourse organization a facilitative category, especially in the circumstance that the reader or hearer would be reluctant or unmotivated to draw together the force of the detail to make the inference-to-purpose the originator intends.

Another facilitative discourse category is what, following Young, Becker, and Pike (1970), will be called a bridge. Its purpose is to inform the reader or hearer about the identity of the topic referent in the event that it might be an unclear concept. The bridge can be realized as a formal definition, a type example, or even as an analogy or metaphor, any rhetorical device which puts the referent into understandable perspective. In conversational analysis it is similar in conception to try-markers (Sacks and Schegloff 1974), digressions from the thread of the conversation whose purpose is to assure that both the speaker and the hearer are on the same wave length.

The type of background information provided in what Grimes (1975) calls the setting of the discourse is functionally

similar to that in the bridge, in other words it activates schema recognition. However, rather than being definitional, its purpose seems to be to embed the topic within a believable or plausible contextual situation. It also serves to delimit the range of variation under which the topic event could be conceptualized. Discourse setting is most commonly found in narrative, although certain forms of exposition, such as historical accounts and biography naturally rely on it also.

The categories mentioned up to this point are those into which the informational content of the message is embedded. Two additional categories have the purpose not so much of specifying the informational base as they do the character and reliability of the communication. The first of these, the register of the discourse (following Halliday and Hasan 1976) reflects the originator's conception of the listening or reading audience. It manifests itself in the choice of apt vocabulary and syntactic structures and reveals itself in an assessment that the style of the discourse is appropriate for the intended audience. In this sense it is a diffuse or pervasive category, rather than one which shows itself primarily at the sentence level. The second category in this type, modality, also tends to be pervasive. Its purpose is to declare or indicate the certainty of the originator's information, for instance whether it is asserted or assumed to be true, whether it is conjectural or hypothetical, whether it is reportedly true, and so on. Both of these categories are evident in narrative and expository discourse, but they probably reveal their complexity more in narrative.

The following paragraph, which was adapted from a longer selection in the CAT, will illustrate the points made above. In the test version of the passage, of course, the sentences would be contiguous and unnumbered.

1. One autumn evening in the early 1800's, twelve-year-old Maria Mitchell sat down and began to adjust an intricate mechanism known as a chronometer.
2. A chronometer is a delicately balanced clock always set at a special time called Greenwich mean time.

3. It determines longitude and, in those days, was vital to plotting and maintaining a ship's course.

4. The chronometer had to be accurate, and its delicate mechanism occasionally required precise technical adjustment.

5. It was for this reason that Maria had been chosen for the job.

The full passage consists of four other paragraphs together making up a brief biography of Maria Mitchell. It is difficult to classify this type of discourse as either narrative or exposition, since in its relation of event sequences it takes on a narrative quality and in its explanation of how these same events shaped the character it resembles an expository account. In any event the typing of the passage is not particularly relevant to the discussion.

The first sentence of the passage states the paragraph topic in the proposition

Maria began to adjust a chronometer.

It also contains setting information in the introductory phrase and the descriptive adjectives. The setting restricts the range of possibilities under which Maria was acting, specifically indicating the event as historical, time-of-day dependent, and within the abilities of a child.

The author also indicates in the first sentence that he expects the reader not to know the word chronometer, thereby preparing the way for the bridge which occurs in sentences 2 and 3. Readers undoubtedly approach a bridge expecting that the definitional terms used in it will be within their vocabulary range. If they are not, the purpose of the bridge in activating a schema will not have been entirely successful. We might expect, then, that since the author indicates the defining term Greenwich mean time itself as requiring explanation, the force of the definition of chronometer is diluted and perhaps ineffectual. The purpose of the second sentence, in that case, may not be realized.

In the fourth sentence the author provides a point of detail within the scope of the topic proposition in anticipation of a question the reader might have concerning why Maria was performing in the way

she was. The point of the detail is, therefore, explanatory and in keeping with the overall intent of the full passage.

The final sentence signals an assessment by the writer by making explicit the reason for introducing the preceding detail. The information rather clearly seems to occupy a separate category from the detail because it conveys the author's intepretation of the situation. In the historical context the information is conclusionary, rather than factual. The assessment also reiterates the topic and in this function serves as a link to the second paragraph of the passage. The second paragraph consists entirely of detail laying out the sequence of steps that Maria engaged in in adjusting the chronometer. It falls fully within the scope of the first paragraph topic and does not--perhaps because of the preceding assessment-- require an explicit restatement of the topic. The reason for including the information within a separate paragraph probably has to do with its focus on Maria's activities, rather than on a characterization of the chronometer.

The register category as indicated above is pervasive and difficult to objectify. It is realized more in the unstated contrast between the structural and lexical choices the author selects to communicate a message and others that might have been selected. What we can loosely say is that the resulting prose takes on a tone which is somewhat academic and aimed at a relatively adept reading group. A rough numerical approximation to the register of this type of prose might be given by applying one or more readability formulas, which purport to measure the difficulty of prose. Our own feeling, however, is that readability formulas have no power to explain the structural basis of register distinctions, and we have not relied on them in this study.

The modality category is also diffusely exemplified in this prose passage. It again shows up more in the contrast of elements than in any specific structure or word. For example, the status of the information presented in the bridge sentences 2 and 3 is different from that presented in the topic or detail sentences. The change in status is signaled by the use of the universal present and equational forms of the verbs in the bridge versus the simple preterite forms in

the topic and detail. The information in the bridge is being declared true by virtue of established convention, while the topic and detail is declared true through an implied attestation. The author is silent regarding the source of the facts in question, leaving the reader to infer that that knowledge is not critical to the development of the account. In other types of discourse, such as accounts of scientific observations, information sources might have to be carefully described. The conclusion stated in the author's assessment signals one more change of status, this one from the preceeding attestation to a probable inference on the author's part.

Coherence and cohesion

At this point it is not possible to formally characterize each of the proposed discourse categories. More study is required before it can be said that they exhaust the full range of possibilities or that their description is adequate. Part of the justification could presumably be derived from studies implicating analogous constructs such as rhetorical predicates (Grimes 1975), semantic macro-structures (van Dijk 1977), or story grammar nodes (Mandler and Johnson 1977). The integration of these various concepts into a single coherent theoretical framework would do much to clarify each of the proposed discourse components. An independently motivated test of the significance of any set of discourse categories would be the demonstration that they are implicated in establishing the coherence of discourse (van Dijk 1977:16), in other words that they make more defined the intuitive judgments we have about what constitutes a well- or ill-formed discourse (Mandler and Johnson 1977:149).

We will take a step to making this demonstration in the examination of test materials in the following chapter, but for now we will propose that the issue of well-formedness of discourse should be approached from two separate directions, one focusing on the coherence of a passage and the other on the cohesion of a passage. The two terms are often taken to be synonomous, but we distinguish them here for the value they have in partialing out the procedures a reader uses in comprehending discourse. Coherence will be defined as that quality of discourse which assures that the informational base

is complete and intact. The presumption that the reading event is a communication requires that all aspects of the message necessary for the reader to approximate the writer's meaning must be present. In terms of the categorical system described above, a coherent account would be one which establishes a topic, defines or specifies referents so that a schema may be activated, provides information in anticipation of questions the reader may have about the topic, makes the purpose of the communication evident, attributes information to its proper source, and does all of the above in a manner appropriate to the reader's language background. A coherent account then is one which is categorically complete.

A cohesive discourse on the other hand is one which conforms to rule systems for combining informational units or categories into syntagmatically acceptable structures. These systems include the rules for producing syntactically well-formed phrases and sentences and rules for linking sentences into well-formed paragraphs. The motivating principles guiding the operation of these rule systems are still only poorly understood (Huggins 1977), although recent descriptive and psycholinguistic studies of cohesion relations (Halliday and Hasan 1976, Clark and Haviland 1977, Chafe 1976), implicate memorial processing as particularly important. Whether information is new or repetitious in the discourse context, whether it must be held for short or long periods of the interaction, whether it requires assimilation with stored information, whether it has high or low saliency are issues whose resolution is described by rules of cohesion.

The cohesiveness and coherence of a passage determine both the availability of information and the ease with which it can be processed. Although each factor contributes to comprehensibility, the relative weight of each will vary with the individual reader. Where one reader might find the information and discourse ties sufficient to draw all the required inferences to activate and manipulate the intended schema, another reader might falter or even be misled. In the next chapter we will examine reading passages and questions in several standardized reading comprehension tests with a mind to understanding the linguistic basis, if any, for why test-takers answer the way they do.

CHAPTER 2

INVESTIGATIVE APPROACH

The investigation was subdivided into two separate but correlated studies: the collection of a set of clinical interviews with children who had taken a standardized achievement test earlier in the school year and the administration of an experimental test. The interviews were designed to probe children's awareness of those points in individual test reading passages which we had hypothesized would interfere with the formation of a coherent and cohesive account. The results of the interview investigation were expected to feed into the experimental test by suggesting particular ways in which test passages could be rewritten to increase comprehensibility. Any positive effects would be ascertained through enhanced performance on rewritten items.

Since a large constellation of subject and test variables enter into the determination of performance we felt it better to maximize our exposure to many of these variables in the interests of tagging those which are relatively more important and eliminating those which were inconsequential. Subject selection and investigative procedures, then, were designed to expose us to a wide mix of variables at the expense of including a large number of subjects in any one cell. It was expected that tests of statistical significance would have to be conservative for this reason.

Subject selection

Subjects were drawn from the third, fifth, and eighth grades of the Montgomery County (Maryland) Public Schools and from two proficiency levels within each grade. The students were selected from these grades since the school district administers the California Achievement Test (CAT) at these levels. Because it would have proven difficult to draw students at both proficiency levels from the same school especially students in the two lower grades, the students were drawn from four different schools: the 3rd and 5th graders from two elementary schools and the 8th graders from a junior high school. The neighborhoods of the four schools were not matched socio-economically and there is good reason to suspect correlation between the proficiency levels of the students and the economic and educational standing of their parents. All four of the schools, however, were participating in

the same experimental reading curriculum so that there was at least a degree of uniformity in their reading instruction.

The proficiency levels were defined as high and low although in fact the low group was actually performing at close to the national norms on the CAT. The initial determination of proficiency was made on the basis of the students' scores on the CAT administered in the Fall of the school year. Students who scored at the 7th, 8th, or 9th stanines on the CAT's reading comprehension subtest were considered high achievers, those who scored at the 4th, 5th, or 6th staines were considered low achievers. Within the lower achieving groups we imposed a second selection criterion, namely that students' scores on the vocabulary subtest of the CAL be as high or higher than their reading comprehension scores. We hoped in this way to eliminate from consideration children whose lack of reading proficiency might be attributable to decoding problems. As a final consideration teachers or administrators familiar with the students were consulted for their impressions of the correctness of a designation. In a few instances they felt that the CAT test scores were unreliable for a particular student, in which case the student was dropped from consideration.

The parents of each potential subject on the approved list were then contacted by mail to elicit approval for their child's participation in the study. It was explained that the child should be appraised of the study and his or her willingness to participate determined. Approximately 70% of the parents responded positively to the request with the majority of the rest not responding at all. Only a few parents voiced their unwillingness to let their child participate. Since most of the non-responding parents were among those of the low achieving students, an attempt was made to contact all of these by phone to give them a more personalized account of the study. This resulted in several more of them granting permission.

We had expected to use 15 students at each of the three grades and in each of the proficiency groups for a total of 90 students. In all but the case of the high achieving 5th graders we were able to reach our required number. For the high 5th graders we made up the difference by drawing the extra students (two) from the school where the low achieving students were selected. This was felt justifiable since no control over situational or environmental variables was attempted. In fact, the students so selected did not vary in any discernible way from their peers in the other school.

All of the students in the final roster were white, were native English speakers (two were bilingual), and were in the appropriate age range for their grade level. All of them had been characterized by their teachers as presenting no significant motivation or attention problems.

Interview procedures

Of the fifteen or more students in each grade-proficiency group, five were randomly selected to be subjects for the interviews and the remainder subjects for the experimental test administration. The only control consideration was to balance the selection of subjects as evenly as possible between boys and girls. The interviews were conducted individually with the author and an assistant. At the start of each interview the students were informed about the study and encouraged to answer questions openly and without anxiety. They were specifically told that they were not being evaluated and that the questioning was for the purpose of determining whether improvement in the tests was warranted. They were encouraged to examine the tests critically and to reflect on their answer choices.

Each interview lasted approximately one hour during which four or five passages from the appropriate grade level of the CAT (Forms C and D) were presented for comment. Some of the passages had been included on the form of the CAT the students had taken in the preceding fall, but few students actually recalled any of the passages and fewer yet indicated that they remembered specific questions. Each reading passage was typed individually on a single sheet of paper. The student read each passage silently after which the original questions asked of the passage were presented one by one on individual 3 x 5 cards. After each answer, the interviewers directed a series of probes at the students, the purpose of which was to discover the assurance the student felt in answering the question, the reason for answering incorrectly if this were the case, and the possibility the student could be directed to see the applicability of another answer. All interviews were recorded and later transcribed.

The interviews were conducted in a clinical manner, meaning that each question on our part was motivated by the particular situation. No attempt was made to standardize the questioning, other than to

assure that the student was looking at each passage question critically. Preparatory to actually interviewing, however, a number of problem types had been isolated in the independent analysis of the test passages and questions. Each of these suggested a particular line of questioning, the thrust of which is suggested by the probes listed under each type. In the examples illustrating each type, the expected answer is starred.

1. Plausible distractors. The organization of the passage suggests that one or more distractors is plausible either because they are coordinate at some level of structural organization or because one is logically included within the scope of the other.

Example: Balloons (CAT 15C)

The passage suggests that the first piloted balloons were used primarily for

*a. sport
b. warfare
c. instrument testing
d. weather forcasting

Both a and b are probable answers, since they are coordinate within the topic continuity: a is cued by the dates 1783 and 1785 and b by the information "before 1800". The difficulty comes with the word "primarily" which requires an interpretation on the reader's part.

Example: Ruth (CAT 13C)

Why was the copy of the map made out of metal?

*a. so that it would last a long time
b. so that it would fit inside a school
c. so that many people could work on it
d. so that the children could take it home

C is a plausible alternative because the passage states that the map is outdoors for people to enjoy. Under an interpretation that the enjoying involves hands on manipulation, then both c and a are logically relatable.

Probes: How sure are you about your answer? Is there another answer that might work here almost as well? Do you see any connection between answers x and y? If someone chose answer x instead, how would you convince him that your answer was better?

2. Difficult inferences. The question asks for causal information leading to or motivating a targeted event, but the information in the passage is laid out in such a way that the motivation can only be inferred from the events following from the targeted event. In a variant form the question requires an inference, for which no information is explicitly stated in the passage. In both cases the reader's expectations about where the information may be located are not realized.

Example: Sequoya (CAT 15C)

Why did the Cherokee nation call a council?

a. to name a new leader
b. to decide where to move
*c. to discuss Sequoya's alphabet
d. to meet with the white settlers

Information pertinent to answering the question unexpectedly follows rather than precedes mention of the target event.

Example: Maria (CAT 18C)

The Mitchells sent the record of their comet observations to Professor Bond because they wanted him to

a. name the comet
b. complete their calculations of the comet
c. tell King Frederick VI about the comet
*d. confirm that no one else had discovered the comet

The passage mentions only that Professor Bond sent the Mitchells good news, making it conjectural what his role was. It could be assumed that his role corresponded to the statements in either b and d.

Probes: How sure are you about your answer? In what part of the story can you find your answer? Do you have to make a connection between different parts of the story to answer the question? Some kids say that y is the right answer. Do you see how they might come to answer that? Do you think that the story is leaving something important out? What should it tell you in order to answer y?

3. Unstated point of view. The question is unclear on whether it wants the test-taker to answer from the point of view of the protagonist or the point of view of the objective observer.

Example: Andrew (CAT 15C)

Which of the following is probably the main reason Andrew Warrel likes the art show so much?

a. He likes animals.
b. He writes for the school newspaper.
*c. He is a member of the Pen and Ink Club.
d. He is a student at Oliver School.

Andrew Warrel himself is presenting evidence to point to d as the correct answer, while an objective analysis would also reveal that he is a member of the club sponsoring the show. The passage itself is presented as a letter with the mention of Warrel's club membership given not in the body itself but in the title following his signature.

Example: Markson (CAT 18C)

Which of the following best describes the surprises in the package being advertised?

a. magical
*b. unknown
c. healthy
d. delightful

From the salesperson's point of view the answer he would like the reader to select is either a or d, whereas an objective assessment of the claims he is making would lead the reader to select b.

Probes: Are you sure about your answer? Did you consider any other answers before you chose your own answer? Would you answer any differently if you didn't believe (the protagonist)? How would you decide if x or y is the best answer? How would you convince someone that your answer is the best one?

4. Hidden detail. The needed point of information is located within a discourse category where the reader would not ordinarily expect it.

Example: Sequoya (CAT 15 C)

Sequoya found that his alphabet

*a. used eighty-six signs
b. took six years to finish
c. would not help his people
d. could not relate to the spoken language

The information asked for is included as a detail within a formal assessment of the paragraph topic. It is consequently completely incidental to the function of the sentence in which it is located.

Probes: Is this information important to the story? Can you point out where your answer is? Is it hard to find where the answer is? Where did you expect the answer to be?

5. Reliance on specialiced background information. The test question requires the reader to bring information to bear which is tangential or irrelevant to the story.

Example: Balloons (CAT 15C)

According to the passage, what effect have scientific discoveries had on ballooning?

*a. Ballooning is safer.
b. Ballooning is more expensive.
c. Ballooning research is restricted.
d. Balloon steering is more accurate.

Readers must understand that individual events mentioned in the story, such as the discovery of helium, constitute a scientific discovery.

Probes: How sure are you about your answer? How did you figure out the answer? Was there enough information in the passage

to answer the question? What do they mean when they ask about (the point of background information)? Do you think this is a fair question?

6. Inadequate identification of referents. The phrasing of the question leads the reader to expect that a referent has been explicitly identified when, in fact, the passage treats the referent as undetermined.

Example: Andrew (Cat 15C)

Which of the following is a fact about "special prize"?

a. Andrew Warrel will receive the prize for writing a good letter.
b. Oliver School won the prize for their decorated balls.
*c. Andrew Warrel wants the club to receive the prize.
d. Some of the drawings won the prize.

The question misleads the reader into thinking some specific prize has been discussed in the passage.

Example: Ruth (CAT 13C)

What did the girl use to get paste out of the bowl?

a. a stick
*b. a knife
c. a brick
d. a paper

The question directs the reader to look for some specifically mentioned girl in the story.

Probes: How easy was it to pick the answer? Did you have any trouble figuring out who they were talking about? What can you tell me about (the inexplicit referent)? Is (the the inexplicit referent) an important part of the story?

7. Vaguely worded questions. The phrasing of the question itself leaves the reader in some doubt about the test-writer's intended meaning.

Example: Alley (CAT 16C)

Which of the following phrases describes the alley as though it relates to a person?

a. "of an old piano"

b. "in a far-off jungle"

*c. "in the back pocket of the city"

d. "on the broad downtown streets"

The verb "relates to" is uninformative in specifying the type of relationship asked for.

Example: Ruth (CAT 13C)

Which word best tells about the boy?

a. lazy

b. silly

c. brave

*d. helpful

The question is stated so broadly that it leaves open the possibility that any of the answers might be justifiable.

Probes: How sure are you about your answer? Do you find the question easy or hard to understand? Is there another answer that might work almost as well? How would you change the question so that someone would find it easier to answer?

8. Absence of superordinate topic. The question suggests a relation between two referents and implies that the relation is mapped in the passage by a statement of topic force which subsumes both; however, the necessary statement is not explicitly stated.

Example: Guitars (CAT 18C)

Today's guitar is most like the guitars made by

a. Zyriab

b. August Otto

c. Hector Berlioz

*d. Antonio Torres

The reader must align passage information within the scope of two separate topics in order to answer the question. The cohesive link of each of these topics to a logically superordinate topic is missing.

Example: Cosby (CAT 13D)

The story says that Bill Cosby has been on TV shows to

a. play football
b. make records
*c. amuse people
d. race go-carts

The question asks for topic level information which the passage itself does not provide. In this particular case the lapse is especially critical because the question can be interpreted to mean either 'why has he been on TV' or 'what has he done on TV'.

Probes: How hard was it to answer this question? How did you figure it out? What kind of connections do you have to make to answer the question? Is all the information where you expected to find it?

9. Passage independent questions. The question can be reliably answered without reading the passage.

Example: Andrew (CAT 15C)

Which of the following phrases is an opinion expressed in the letter?

a. "I am writing..."
b. "It will be shown..."
c. "It's the best show..."
d. "Three are ink drawings..."

Assuming that a reader has a clear idea of what an opinion is, the question can be answered without consulting the passage.

Probes: How easy was it to answer the question? Can you find the answer in a particular part of the story? If that part of the story was missing do you think that you could still answer the question? Why do you think that this question was so easy?

10. No best answer. The information in the passage suggests as the most plausible answer an option which is not included in the question.

Example: Sequoya (CAT 15C)

Which of the following best describes Sequoya?

a. lazy
b. lucky
c. selfish
*d. intelligent

The passage provides no direct information which would support one of these answers over another. The best answer, which would be persistent or concerned, is not included.

> Probes: How satisfied are you about the choice you made? Do you think there is some other choice which should have been included? If someone said that none of the choices was good what would you say?

Following the substantive part of the interviews, and depending on available time, students were questioned about their feelings regarding testing, their attitudes toward reading, and their sense of how their reading behavior shaped itself to the situation of test taking.

All of the students without exception were cooperative during the interviews, although they spread themselves over a wide range of variability on their willingness to initiate discussion, to volunteer information, and to test various options. The analysis of the interviews is presented in the following chapter.

Experimental procedures

The experimental procedure was designed to test the hypothesis that reading passages from three standardized reading comprehension tests--The Iowa Test of Basic Skills, the California Achievement Test, and the Metropolitan Achievement Test could be rewritten, such that they were more easily comprehensible. The measure of comprehension was performance on the same set of questions which test-takers were required to answer on the original versions.

The complex considerations which entered into making individual changes are more fully described in Chapter 4. At this point it will suffice to say that the rewriting resulted in no significant changes in the determined readability of any of the passages, that no new information was deleted. Essentially all changes were made in consideration

of the theory of discourse structure laid out in the preceding chapter. Other than to attempt a consistency in the formatting of passage information, no other control over the type of change made was possible since each passage selected for rewriting presented unique problems of discourse organization. Furthermore, the aggregate of the passages selected for any one grade level demonstrated a wide variety of discourse types within the broad category of expository prose. The purpose was to select items which in their totality represented the full range of stated test objectives.

Individual test packets were constructed for each student and the packets were mixed for original and rewritten passages. Third grade students read four passages altogether, 5th graders and 8th graders each five passages. Students, of course, were not informed about whether a particular passage they were reading was original or rewritten. Each of the packets was designed to be easily completed within 45 minutes, and, in fact, no student was unable to complete the test in the allotted time.

The test was in most respects administered under conditions similar to those which would have obtained under a real testing, except that a machine scorable form was not used. Students were, instead, asked to indicate their choice in an appropriate blank. In addition, they were asked, for each question, to provide a confidence rating for their selection. The point of asking for this information was to focus the student's attention to the task, not to provide another scoring variable. Consequently, except as it was useful in disallowing the data from one student, the confidence ratings are not further discussed.

A total of 63 students were administered the experimental test with the breakdown by grade and proficiency level as indicated in Figure 2-1.

Figure 2-1

Subject population on experimental test

	Low Achieving	High Achieving
3rd	11	10
5th	10	11
8th	11	10

In addition one other low achieving 3rd grader took the test but his scores were not tabulated since there was reason to believe he guessed at all answers.

Following the test administration we attempted some informal interviews with students in group situations to elicit their feelings about the relative difficulty of individual items. It was hoped that students' opinions might systematically vary as a condition of which version of a passage was read. Unfortunately, time constraints and logistit difficulties made it impossible to approach the end of exam interviews with any precision and they are not further considered.

CHAPTER 3

INTERVIEWS

The interview procedure was initially conceptualized as a fact-finding study to identify the characteristics of reading comprehension tests which perplex children at various ages. It was expected that the types of structural problems we had identified in the tests would likely prove difficult to children, but that they would be able to vocalize their difficulties and suggest improvements. We then expected to implement their suggestions in designing the rewrites of individual test passages, and to verify the effect of the characteristics in an experimental test.

What in fact resulted was much less straightforward. Students as a whole did find certain of the hypothesized problem types difficult, but others they found to be relatively untroublesome. With certain of the difficult types, the subjects sorted themselves by proficiency groups, with the lower achieving subjects experiencing essentially all the problems. With certain other types, however, students at both proficiency levels professed difficulty, but the higher achieving students produced proportionately fewer errors. This skewed pattern of results suggests that the ten identified problem types may not be equatable on the same parameters and that a finer analysis is necessary in order to understand the mechanisms underlying judgments of difficulty and poor test performance. We will suggest in the following discussion that factors other than those which fall under the usual set of reading comprehension objectives are in part responsible for perceived and actual item difficulty. The results will show that difficulty is often attributable to not knowing the appropriate procedure for verifying an answer or for selecting a comprehension strategy. These factors in turn may be consequences of a student's negative attitude toward tests or to his own test-taking skills. The distinctions between each of these factors and the skills encompassed in "pure comprehension" are weak, allowing for some conceptual overlap. In the following discussion, then, we will not attempt to rigidly define each of these categories other than to characterize them as metacomprehension and test-wiseness skills.

The claim that we will make from the interview results is that reading tests, in addition to their explicit purpose of assessing comprehension skills, also covertly assess metacomprehension and test-taking skills. To the extent that test developers do not recognize this point or are ignorant of the interrelations between these various aspects, the breakdown of skills which they claim a particular item measures is likely to be erroneous or only partially correct. It would be simplistic to claim, then, that a low-scoring student may need remediation in some comprehension skill, such as drawing inferences, without first ascertaining that the fault does not lie in some other direction. In the discussion we will concentrate first on detailing the variety of extra-comprehension factors which influence test performance and later consider the effect that poorly written passages and test questions can have on deflecting readers.

Comprehension Strategies

Quite early into the interviews it became apparent that both high and low achieving students were operating with the same repertoire of comprehension skills, if we restrict the term to the usual array of literal, inferential, and evaluative skills. The high achievers, however, were more successful in consistently applying these skills to the solution of particular problems, while the low achievers almost inexplicably would one time apply a needed skill and another time not. What was particularly troubling was that an individual student's performance could not be predicted for any particular item. Except for a few items with inherent difficulties in their construction which posed problems for all students, no item initially judged to be without structural inaccuracies was invariably failed. Contrary to what one might predict about a comprehension skills gap between high and low achieving students, there appeared to be only a difference in the frequency with which particular skills were applied.

Questioning, however, did eventually make it clear that an absolute difference could be found to explain the scatter in the response pattern. Rather than involving a traditional comprehension skill, though, it involved the selection of a particular processing routine: either a hierarchical processing strategy, which represents

the preferred algorithm for reaching the correct answer, or what we will call a detail processing strategy, in which all information is essentially equated for relevance. It is probably the case that the two processing models can operate in tandem or that one can supplant the other for some period during the reading event, but in general, for any particular passage, students seem to prefer one to the other.

The distinction is clarified in the following two exchanges, each dealing with the same question on the Guitar passage (CAT 18C). The question reads

> Which of the following is a guitar-like instrument from Asia Minor?
>
> a. lyre
> b. nefer
> c. cithara
> d. vihuela

The passage itself mainly concerns the history of the guitar, and as the following segment shows the information relevant to answering the question is incidental to the theme.

> One of the earliest instruments shaped like today's guitar was the nefer of ancient Egypt. Later, other instruments related to the guitar were developed. Among these were the vihuela from Spain and the cithara from Asia Minor.

The first student is a low achieving 8th grade female reader (CK) who nevertheless successfully answered most of the questions on the passage, including this question. The interview begins by asking for her opinion of the importance of this question. (The letter I indicates the interviewer and the letter S the student subject.)

> I: Do you think what they're asking you for is an important part of the story?
>
> S: No.
>
> I: If it's not an important part of the story, how did you know where to find it?
>
> S: It'd probably be in the middle.

I: And how would you go about finding it, if you didn't remember it from reading the story through the first time?

S: Look through it real quick for the word.

I: Which word would you look for?

S: Look for the first word (i.e., the first option).

I: Do you know any of these other words (i.e., the other options)? Have you ever heard of them before?

S: I've heard of that one (option d). It's a musical instrument. (Later it became evident that she had understood vihuela to mean viola.)

I: Did you think this question was fair?

S: Yeah.

Later in the session and relative to another question she clarified her conception of fairness.

I: I want you to think about whether or not this is a fair question.

S: No, because they don't say it in the story....they make a hint, but it doesn't match up with any of them (i.e., the answer options).

It is clear from the larger context that she means none of the answer options is clearly indicated in the passage, either overtly or through a meaning link. These excerpts indicate a preference for the detail processing strategy. Information is evaluated not so much for the role it plays in a paragraph, but for its own sake.

In the second dialogue a high achieving 8th grade male (3A) implicates quite a different processing strategy, one in which hierarchical relations become heuristically important.

I: Have you ever heard of these instruments before?

S: Yeah, I've heard of the lyre.

I: The other three possibilities (i.e., the other answer options)?

S: No, I haven't heard of any other than the lyre.

I: Does that at all make it difficult to locate something if you don't know what it is?

S: Not really, because it says in the passage that they're guitar-like instruments and I don't expect to know all the instruments.

I: Just the very fact that it's a word that you've never seen before, does that make it any more difficult for you?

S: No, that doesn't make it any more difficult ....You can tell they're all musical instruments. You don't need to really know the definition.

I: Is this question asking for an important piece of information?

S: No, I don't think it's really that important. It's critical reading I guess, to pick that out. Knowing that the cithara came from Asia Minor is not what they're trying to get across.

Notice that both students have accurately assessed the relative importance of information asked for, indicating that each has a clear conception of the passage theme and probably also the organization of discourse categories in the paragraph. Each also agrees on the merit of the question although the high achieving student mislabels the demand as "critical reading." The difference between the two lies in the conception each has of the information per se, the low achiever seeing it as coequal to other passage information, the high achiever as nested within a more encompassing detail. As we have seen in this example, it is a difference which may not reveal itself in a difference in response, but only in a difference in the procedure used to verify the selected answer.

Each student found it necessary to recheck the passage before answering, but each went about it in a different way. The low achieving student systematically scanned for each of the alternatives, while the high achieving student presumably first located the superordinate concept from his memory of the organizational hierarchy and then verified his choice. Both procedures involve a process of elimination, but the process operates at two different levels of analysis, the one making no assumptions about the location or the connectedness

of the options, the other making an assumption that the options are associated semantically in an inclusion relation and that they will most likely be found in an equivalent location and in the same discourse function within a paragraph.

Each of our assertions, of course, is conjectural, but each holds up through repeated instances of asking students to locate information. The detail processing strategy reveals itself in scanning behavior; the hierarchical processing strategy reveals itself in a directed search to verify some hypothesis. Either approach can be successful and either can lead to an incorrect choice, but overall, the hierarchical processing strategy leads to more correct answers, if only because certain questions predispose themselves to this kind of analysis and because it saves time. It is, however, a strategy which many low achieving readers prefer to avoid probably because it reduces the certainty of response. Many of these students know that detail processing will be successful in instances where the information logically falls within discourse detail. Hierarchical processing risks seeing this information as trivial and relegating it no status in memory. If the passage information should be queried, as it often is in comprehension tests, hierarchizing readers may not have access to it. The detail processing strategy at least assures those who select it that they will answer correctly as many questions as they expect to answer beforehand. As the low achieving student above put it in response to another question

I know I'm going to get at least one wrong.

Outcomes of Detail Processing Strategies

A reader's decision to rely on a detail processing strategy should produce some success with the types of comprehension questions which require a direct or easily inferrable match-up between passage content and answer option, but it can lead to difficulties in processing questions which depend on an awareness of the hierarchical relationships between information units or which require a tracking of information flow. In the first case it may happen that the reader fails to apprehend the main idea of a passage or, more commonly, loses sight of the relative importance of information. For instance, a low achieving 5th grade female student (MY) responded in the following way

to inquiry about the following question from the Balloon passage (CAT 15C)

In what year did Letitia Sage fly in a balloon?

The relevant passage information is contained in a discourse detail.

Among the first adventurers to try ballooning was Letitia Sage, who flew in a balloon in England in 1785.

The student was not being successful in locating the information and after some time the interviewer began the dialogue.

I: Do you think that what they're asking for is important to the story?

S: Yes.

I: How do you mean that it's important to the story?

S: Because right now we're working on reading comprehension. And h[illegible] would they know that you really read the story?

I: O.K., so it's important if somebody's going to ask you about it. Is it important if you were just going to read it for your own pleasure? Would you remember that sort of thing?

S: I would tend to remember it. In case my mom starts asking me questions about the story; [illegible]u know, did I have any problems with the story, and stuff like that. So then I get teaching at school and at home, 'cause of my mom and dad.

The student is evidently basing her selection of a comprehension strategy on the importance that others attach to content, not the importance that the information might take given a logical analysis or a highly felt personal motivation. She is, in fact, not reading for herself; she is reading for someone else.

Tracking difficulties are particularly evident among the younger students. For example, a 3rd grade low achieving female (AC) correctly answered the following question from the Cosby passage (CAT 13D) with the fourth alternative

When he was a boy, Bill Cosby

a. Made records
b. became famous
c. appeared on TV
d. made people laugh

but through a more involved procedure than the text warranted. The relevant part of the passage read

> Can you tell stories that make people laugh? Even when he was a boy, Bill Cosby was able to do this. His teachers and friends enjoyed his stories.

As is almost obvious to an adult reader, the answer is found by making an inference between the first and second sentences. The young reader in question, however, failed to spot the connectedness of this information and opted for a processing routine in which the deictic word this ending the second sentence was interpreted as pointing in a forward rather than backward direction. This was revealed in the dialogue.

I: Are you having trouble (answering the question)?
S: I can't decide.
I: Which ones are you thinking about?
S: 'Made people laugh' because he is funny now, so he had to be funny when he was little.
I: Is there any part of the story where it actually told you he made people laugh when he was a boy?
S: No.
I: It doesn't say that?
S: No. It says 'Can you tell stories that make people laugh? Even when he was a boy, Bill Cosby was able to do this. His teachers and friends enjoyed his stories.' So he's got to be funny if they enjoyed his stories.
I: Oh, I see. What does 'this' mean? (Interviewer pointed to the occurrence of the word in the second sentence.)
S: It means he was able to make people laugh.

I: When you see this word 'this' does it mean what's going to follow it?

S: Like, Bill Cosby was able to do this: his teachers... (Emphasis in the original.)

We would contend that the child is not suffering from some inability to make inferences, since she does successfully infer an answer from the evidence in the third sentence. She is processing material in each sentence as if it were essentially independent except when a specific cue like the deictic word indicates an explicit connection.

The same student gives what may be another example of the effect a detail processing strategy can have on comprehension. In another question from the same passage the reader is asked

The story says that Bill Cosby has been on TV shows to

a. play football
b. make records
c. amuse people
d. race go-carts

The relevant passage information is spread over two paragraphs.

His records and TV shows have made him one of the best-liked and most famous people of his time.

In his shows, Bill Cosby often talks about the things he did when he was a boy.... He tells stories about playing football in the street and racing go-carts.

The student selected the fourth option because

S: ...I remember they said in the story 'go-carts' somewhere or other.

The dialogue continues

I: O.K., but here they say about playing football, too, don't they? Maybe that's the right answer then?

S: O.K., let's see: 'playing football in the street and racing go-carts'.

I: Did he race go-carts when he was on TV?

S: No, it doesn't say that.

I: Would you still answer it that way, do you think? Even though it didn't say that?

S: No. I don't know.

I: Does the question ask you what he did on TV or why he went on TV?

S: The first one: What he did on TV.

I: Does it talk about being on TV at all in the story?

S: It says 'His records and TV shows have made him one of the best liked and most famous people of his time.

I: O.K., does that tell you why he went on TV?

S: No.

I: So we're back to the start then. How're you going to answer this question?

S: I have no idea.

What has evidently happened in this student's reading of the question is that she assumes it to refer to an explicit activity which Bill Cosby performs on TV, when in fact the intent of the question is to determine the effect of his performance. The question itself permits either interpretation, but the second one is almost precluded if the processing strategy does not allow for a linkage between the previous mention of making people laugh and telling stories on TV. Although we did not extend the interview, we would suspect that the child would have preferred and immediately chosen an option which read to 'tell stories.' To her mind, then, there was probably no correct/answer among the options. Also, as the questioning shows, she had no appreciable difficulty in comprehending the passage using her set of preferred procedures.

Verificational Routines

The interviews also revealed a significant difference between high and low achieving students in the procedures they used to verify an answer once one had been selected. In general the high achieving students were more concerned about verifying information, and they set about it in a more principled way: for unambiguous questions

simply finding as close an equivalence between passage content and item query as possible and for ambiguous questions adopting a process of elimination procedure. For instance, in the following comment a 5th grade high achiever (CM) describes her preferred routine.

> See, lots of times in these questions that they have on the test, they don't say which one is not this. Some of the questions you have to look through and find the one that is most reasonable. Lots of times, none of them are, but you have to just look really closely.

An 8th grader (SG) proffers a similar opinion in dialogue regarding an item that she answered assuredly and without error.

> I: What was the strategy that you used?
>
> S: I just eliminated. It couldn't have been how far it was from Polaris because it never even mentioned that; and notify the U.S. Coast and Geodetic Survey. That's where her father worked, and didn't have anything to do with it; and send news to King Frederick--they didn't send news, he found out himself.
>
> I: Now, based on the discussion that we've just had about this question, do you consider it to be fair?
>
> S: It's not unfair.
>
> I: Would you predict that somebody who didn't read as well as you did would have a problem with it?
>
> S: I suppose they would. But if they just use their common sense and read the whole thing...
>
> I: By common sense, you mean going through each of the possibilities and ticking them off?
>
> S: Yes.

Quite often a request to verify an answer led a high achieving student through the justification of a process of elimination. By contrast, the same request of a low achieving student provoked a quite different response. For instance, a 5th grade low achieving male (GJ) correctly answered the following question from the Novelty passage

(CAT 18C) by selecting the 4th alternative.

> Why does the speaker ask if life has 'robbed you' or is 'getting you down'?
>
> a. to ridicule the listener's feelings
> b. to inform the listener of dangers to avoid
> c. to persuade the listener that no risk is involved
> d. to suggest that the listener is missing out on something

The expected answer is cued most importantly by the following passage sentence.

> If it (life) is (getting you down), we can offer you a way to come alive by putting a surprise in your life.

The interview began with a request to verify the answer.

I: Why are you thinking that it's 'd' and not one of the others?

S: Cause it's asking you if life is getting you down, to make you think of all the bad things of life, and try to persuade you to get this product, and that will keep you from getting robbed.

I: What about "b', since you said "the bad things in life?"

S: I thought that the salesperson is trying to keep you from buying anything else.

I: You mean like buy another product instead? Well, what about 'c' then. Maybe that makes 'c' a better answer?

S: Oh, yeah. 'c' looks good. 'c' is probably better.

I: Now explain to me why you think 'c' is better.

S: Cause he's trying to persuade the person.

I: O.K., but does he actually say something about no risk involved? What kind of risk do you think he's talking about?

S: It says 'getting robbed' and a lot of people go out and buy stuff and they get robbed.

I: Oh, I see. And he's saying that this wouldn't happen in this case, so there's no risk.

S: Yeah.

The change in position which the interviewer was able to negotiate in this and many other cases indicates that students can be confused rather than helped by attempting to verify their answers. In this case the student had not grasped the figurative reference of life robbing someone and had used his own conception of it to rationalize another answer.

While it was possible to redirect some low achieving students to reconsider their first choices, other students with the same reluctance to verify expressed their unwillingness in a refusal to consider other alternatives, although it was often clear that they suspected options other than the ones they chose might be correct. The following dialogue /with/ a third grade low achieving girl demonstrates this point.

I: Does one of them seem like it might be better? How would you decide? If somebody says, "I like this one," and somebody else says "Oh, I like this one," would you have a fight about it or what?

S: That person goes with his, and I go with mine.

I: But only one can be right, remember.

S: So I'd find out which one is right and which one is wrong.

I: What if you were wrong? Then what would you say?

S: I wouldn't say anything. I'd just say I was wrong.

It is certainly not assurance which motivates the obduracy of this student's position, it is resignation to a belief that tests are somewhat capricious. Students in this category feel incapable of resolving the conflict inherent in a set of answer options in a way which will invariably lead them to the expected answer, and so they give their allegiance to the answer which fits best with the schema they have

elaborated for comprehending the story. To their credit, it rarely happens that they are unable to provide some plausible rationale for their choice, just as GJ did in the example cited earlier. Answers given rarely seem to be without some basis in fact or imagination.

Schema Activation

The difficulty that some low achieving readers have in verifying their selected answers marks the lack of a metacomprehension skill, rather than a comprehension skill per se. The low achieving readers we studied, in fact, gave every indication that they invoked the same generalized reading strategy--invoking, activating, and reformulating an underlying schema--as did the high achievers. A difference did show itself, however, in regard to how much dependency was accorded the underlying schema. Low achieving students generally seemed less willing to foresake or drastically alter their schema in favor of investing in a new schema, while the high achieving students typically showed no such reluctance. This pattern of results is consonant with our other findings particularly the difference in processing strategies discussed earlier. A reader who approaches new material one sentence at a time is forced to assess that information in relation to some underlying conception in order to reach conclusions which relate new information to itself. In other words, the integration of information which is presumed to take place during hierarchical processing is achieved during detail processing through the mediation of an underlying schema. If the schema is accurate and full, then the new information presumably finds its properly integrated spot without the type of reading implied by hierarchical processing; in which each statement is evaluated first for structural relationship to other statements and later for its fit to a preexisting schema.

Possibly either procedure can produce full comprehension, although the methods for demonstrating comprehension would have to be sensitive to the procedure used. However, if the reader's existing schema is only weakly developed, comprehension will be more difficult for detail processors, since their fit of the new data is more likely to be erroneous. Evidence of this claim can be found in many interviews and at all grades, but it is especially evident with the youngest subjects.

For instance, a 3rd grader had difficulty answering the following question from the Ruth passage (CAT 13C).

> Why was the copy of the map made out of metal?

She began the discussion of the item by remarking that the passage didn't tell why--which, in fact, it does not. It states

> Two years later Ruth Asawa made a lasting metal copy of the paste map.

making the reason dependent on interpreting the word 'lasting' as inferentially related to the cause. During the interview the student on being asked why she though Ruth used metal for her map immediately and without consulting the passage selected the expected answer.

> So that it would last a long time.

The interview then continued from this point.

I: Explain to me why that (choice) would work?

S: Because if there was a fire in the school, the thing would just get hot, it wouldn't burn or anything like that.

I: So it would last a long time. Could it be like the third (choice) - so that many people could work on it? How would that work out?

S: If many people had a whole bunch of equipment, if they drop stuff on it, it probably would dent, but it wouldn't ruin the map.

I: O.K., so you can figure out a reason for that one too. Which one do you think is the best one?

S: I think the third one (i.e., people working on it).

I: Remember where they talked about the metal map in the story? Was it the beginning, middle, end?

S: I think it was in the middle.

(Student had a great deal of difficulty locating the reference using her scanning procedure, but after some period of time finally found it.)

S: Oh, here it is. Last paragraph. It says, 'Two years later Ruth Asawa made a lasting metal copy of the paste map.'

I: What do you think is going to be the right answer?

S: It's going to be the first one (i.e. last a long time)

I: Why?

S: Because it says 'lasting' - lasting a long time.

This interview, besides indicating the student's reliance on her own background information in answering the question, illustrates several concomitants of adopting this approach. First, she is unable to locate information based on her understanding of the story events, although she has a clear recollection of the mention of the metal map. Second, in the absence of an overtly stated reason for making the map out of metal she contrives one reason, then another, and weighs the probability of each being the preferred choice against criteria external to the story. In first selecting the correct answer, notice that her reason is peripheral to the evidence as presented in the passage, which, in effect, means she has selected it, if not entirely by chance, then for the wrong reason. As an indication of her true passage comprehension, then, the question is uninformative. Third, she imports information from her underlying schema into her conception of the story, making reference to the possibility of a fire, to some unspecified equipment, and to the setting of the events in a school. Finally, she rejects all of the aforesaid logic in favor of an explicitly stated reason which intrudes itself during a verificational procedure.

The procedures which this and other students have followed in reading test passages do not necessarily comprise a poor comprehension strategy. To the extent that they involve the student in restructuring a schema they perform just the task they should as a reading event. However, to the extent they deflect the student from selecting the expected answer on a test item, they fail to achieve their specific purpose. If these suggestions can be independently verified, then it would seem that reading comprehension tests are either biased toward the hierarchical processing strategy or are measuring non-

comprehension tasks. In either case the student who has successfully learned to cope with this form of instrument would be at an advantage in comparison to one who has not.

Test-taking skills

The analysis of the interview data provided several independent examples that, comprehension issues aside, the CAT was also measuring test-taking skills. Certain of these skills, of course, are necessarily involved in taking this or any form of structured test, and to a certain extent our educational system has subtlely or overtly prepared students to develop them from an early age. The comprehension tests recognize the contribution that test-taking ability can make in affecting a student's score, but they maintain that these skills are sufficiently taught and learned by means of the instructions to the test and by the administration of practice problems accompanying the test. Our discussion here, however, will exemplify the kinds of skills which are rarely taught, but which must be grasped if students are to succeed on tests.

In a common type of problem a test item sets up the possibility of two logically correct options, one of which is literally verifiable, the other inferentially verifiable. In this situation there is an unwritten and untaught rule that the literal answer is to be preferred over the inferential and should be selected. Students when confronted with this kind of choice either recognize the difficulty and weigh the alternatives or fail to recognize it and go with one or the other based on their understanding of the passage.

A particularly troubling item was included on the Plants passage (CAT 15D). It read

Why does the sundew move its tiny red "arms"?

a. to catch insects
b. to spread a sweet smell
c. to feed new young plants
d. to gain an insect's attention

The passage literally supports only the first option through a link with the passage theme, but the fourth alternative is inferentially possible also. One 5th grade high achieving student (CM) immediately noticed the difficulty.

S: 'A' and 'd'... It says 'It has red vines that look like arms, and these vines are covered with sticky drops.' And it could be to catch the insect or it could be to gain its attention...Well, 'a' or 'd,' but most likely when they're talking about trapping insects (as) the main thing of this (passage), 'a' would be it. But if it wasn't (the main idea), if it was just like a story that didn't say really any specific topic, then it could be 'd.'

I: Then you would answer 'a' on the test?

S: They're both tricky. I don't know which one right now. 'D' seems to be more what I'd pick, because I don't think the sundew waves its arms to catch insects. I think that's just its natural way of doing things.... But moving its arms isn't what catches it, it's what attracts the insect to come near.

A second, low achieving 5th grader (MT) also saw the dilemma, but focused on 'b' as a possibility, because he assumed the sticky drops on the sundew's arms exuded a smell. When alerted to option 'd', he also admitted its possibility. Unlike the preceding student, however, he had no reasoned procedure to guide him out of the problem and admitted finally that he would probably mark two answers correct.

A second type of commonly encountered test-taking difficulty concerned CAT passages which presented a position advocated by some person other than an unspecified narrator. Certain items based on these passages asked the reader's judgment regarding a particular claim made by the advocate, but in doing so did not specify whether the reader was to judge the claim from the standpoint of the advocate or the objective bystander. The Andrew passage (CAT 15C) provides a good example. In one item the reader is asked,

Who said that "Creature in Ink" is the best display the school has ever had?

a. Andrew Warrel
b. everyone who saw the show
c. the editor of the Oliver News
d. all the students at Oliver School

The passage provides the following information.

> The show is titled "Creatures in Ink."...It's the best show Oliver School has ever had. Everyone who sees it will agree.

One high achieving 5th grader (MB) responded to the question as follows.

> S: Well, I'd say this is 'a,' even though it could be a lot of the other ones...Because you can tell from the children's enthusiasm that they think everyone who saw the show would think it's the best. And all of the students at Oliver School think it's the best. (But) Andrew Warrel is the only one who really says it.
>
> I: How difficult would you say this (question) is?
>
> S: This one is difficult in a new way. I've seen a few but not very many that are difficult in this way, where you could have a lot of possible answers.

A low achieving 5th grader (MY) initially chose the expected answer but could be dissuaded from it to the other option.

> S: 'A', Andrew Warrel.
>
> I: Are you sure its not 'd' (i.e. all the students at Oliver School)?
>
> S: Oh, yeah, 'd' because all the students are saying it's good.
>
> I: Originally you said it was Andrew Warrel. Why did you think it was Andrew Warrel to begin with?
>
> S: Because Andrew Warrel is writing to the editor.

Both of these students perceived the difficulty in answering the question without a declared point of view, but whereas the high achieving student resolved the difficulty in favor of the objective bystander position the low achieving student resolved it in favor of the advocate's claim. We might have predicted this pattern of behavior given the predisposition of the low achievers to stick as closely as possible to the stated message. Still, it is quite evident that the probability of the expected answer had not escaped the low achiever,

meaning that her 'wrong' answer cannot be construed as a failure to comprehend the passage, only as a failure to ascertain the test maker's intention.

Hypothesized problem type

The opening remarks to this chapter mentioned that the pattern of responses to the ten problem types identified in the preceding chapter was uneven, that some of the types did not produce the hypothesized difficulty for any identified group and others were not consistently difficult for any single individual. Furthermore, of the types which did prove difficult the responses showed that the burden of the difficulty was borne by the low achieving students. The high achieving readers, while often noting the potential for confusion and error, nevertheless claimed they had the skills necessary to avoid answering incorrectly.

The status of each hypothesized category as an actual problem type is provided in Figure 3-1.

Figure 3-1

Hypothesized problem types and determined status as an actual problem

	Unsubstantiated	Verified
1. plausible distractor		x
2. difficult inference		x
3. unstated point of view		x
4. hidden detail	x	
5. reliance on background information	x	
6. inadequate identification of referents		x
7. vaguely worded questions	x	
8. absence of superordinate topic		x
9. passage independent questions	x	
10. no best answer	x	
11. unknown vocabulary in question		x

The status is listed as either verified or unsubstantiated. A verified problem is one which consistently proved to be difficult for students or which provoked substantial discussion. An unsubstantiated problem is one which, contrary to expectation, proved to be easily answerable or to arouse no recognition of difficulty. In addition to the ten

hypothesized problem types discussed in the preceding chapter, Figure 3-1 also includes reference to one additional problem type which was unanticipated but recognized and verified during the interviews. In the following discussion the unsubstantiated problem types will be discussed first, followed by consideration of the verified problems.

Unsubstantiated problem types

Hidden detail. None of the CAT questions identified as involving a hidden detail, including the question from the Sequoya passage exemplifying this type (p. 31), proved at all troublesome to any of the students interviewed. Moreover, attempts to point out the hypothesized difficulty of such questions were largely unsuccessful. Students seemed to elevate the status of the queried information to accord it the actual importance of its position within the passage. For instance, the Sequoya passage question places what we imputed to be an incidental detail--that Sequoya's written language contained 86 letters--in the elevated position of a passage assessment. One high achieving 5th grader's (MB) response to this out-of-context mention was to give more weight to the information. In response to a question asking why he thought the information important at all, he said

> S: It's always important to know the amount, and 86 signs tells you that he must have worked hard enough to really do a thorough job of the alphabet.

What he had done was to assign an importance to the information consistent with the discourse category in which it was included.

It is not as easy to ascribe the same logic to low achieving readers, who generally were not vocal about answers they were assured of. However, given their inclination to detail processing it is possible that the fact of 86 letters would be precisely the sort of information which would capture their attention on its own right. In fact, several of these students answered this question without any look-back.

Reliance on specialized background information. This category proved to be insufficiently defined to distinguish it from other verifiably difficult or unsubstantiated problem types: In certain instances the background information required of the reader was sufficiently generalized that all of the students interviewed knew the

information--for instance, the information that a grapefruit is a fresh fruit. The problem of an imprecise definition revealed itself especially in confusions with unstated point of view category, as for instance, in the question from the Andrew passage discussed above (p. 30)

> Which of the following is probably the main reason Andrew Warrel likes the art show so much.

Originally this was categorized as requiring the specialized background information that writers may not be entirely objective in advocating a particular position or opinion. The correct answer, which asserts the reason to be Andrew's membership in the club sponsoring the art show, however, was later felt to be better characterized as a failure on the part of the test developer to clearly specify whose point of view the student was to take in answering the question.

A second merging of criteria was found between this category and the category of hidden detail precisely because the information necessary to answer the question was so diffusely stated that the reader was as likely to reach a decision on the basis of background knowledge as on the basis of information presented in the passage. For instance, a question in the Ellen passage (CAT 13D) asks the reader for the meaning of the word 'nearsighted' the information for which is not explicitly given in the passage at the point where it would most likely be expected, prompting a conclusion that the question required background knowledge.

Vaguely worded questions. This category, like the preceding, is also inconsistently defined. In certain cases, such as the question from the Ruth passage discussed above (p. 33)

> Which word best tells about the boy?

most students were able to retrieve the correct answer. In other cases, however, a vaguely worded question did prove troublesome, but this may have been a consequence of the intrinsic difficulty of the information content of the question rather than a factor of its structural form. In general, it seems that when faced with a question categorized as vague, the student is forced into a process of elimination strategy during which one of the options tends to stand out more prominently than the others. The student will select this option, but will oftentimes express some doubt or uncertainty about

the choice. These students will also sometimes express their dissatisfaction with the fairness of the question. The following interview with a low achieving 3rd grader (EB) on a vaguely worded question from the Cosby passage (CAT 13D) illustrates these points. The question read:

> Bill Cosby's stories show that he probably was
>
> a. happy about being young.
> b. excited about watching TV.
> c. worried about his brothers.
> d. curious about making records.

The student initially found it hard to pinpoint an answer and the interview followed him through his testing of the various options. At a point when the difficulty of the question was entirely evident to both the student and the interviewer, the interviewer asked

> I. What makes this one so hard to answer?
> S: I don't know.
> I: It's a hard question for you, then, right?
> S: Yeah.
> I: Is it a fair question?
> S: Yeah.
> I: Why is it fair, if its so hard?
> S: If you read the story, you should know the answer.
> I: You did read the story, and you've been telling us answers all along. How come this one is so hard, so much harder?
> S: By the time of the last question, you always forget the story.
> I: Maybe that's a good reason. What about if you just guessed? Which one would you guess.
> S: Probably 'a' (happy about being young).
> I: Why 'a'?
> S: 'His stories helped everyone remember what it was like to be young.'
> I: And that comes pretty close?
> S: Yeah.

I: Is it a guess or is it for sure?

S: It's a guess.

I: Do you think it's fair for them to ask you to make guesses like that?

S: No, because you might get it wrong, right, wrong, right..

Passage independent questions. Students interviewed only rarely regarded test questions as being passage independent, at least until the possibility was specifically addressed during the interview. Even in this case, however, students almost overwhelmingly felt that the selected answer should be verified, either through lookback to the passage in the case of the better readers or through some logically defensible connection to passage information in the case of the poorer readers. In any event, given the experimental procedure, all students did in fact read the passage before they were given the question to answer. This would preclude the possibility of reading and answering a question totally independently of the passage. The points made are illustrated in an interview with a high achieving 5th grader (KE) on the question from the Andrew passage exemplifying this category (p. 34).

I: What about if you didn't read the story at all. Could you still answer it?

S: Probably. But it's better to read the article so you know what you're talking about. But you could do it without reading it.

I: Did you answer without reading it (that is, without re-reading it).

S: Yes.

I: Is it still a reading comprehension question, if you don't have to read the passage to answer?

S: Well, it's not really reading comprehension, but it could be put there because it is about reading, and there's not quite another place to put it.

No best answer. Students tended invariably to approach questions which had been categorized as not including the best answer among the options as if they were simply more difficult questions in which the

best answer among the options was elusive. At most, students could be led to seeing the questions as unfairly asking for distinctions which were too finely drawn, a consequence of none of the options being directly on target with the information in the passage. In situations such as this, the low achieving students especially tended to fabricate the justification for selecting the answer they chose. More often than not, however these students were able through their comprehensive understanding of the passage to select the item which most closely approximated what the best answer should have been. Consequently, few students tended to answer these questions incorrectly for the hypothesized reason.

Verified problem types

Plausible distractor. The category of questions for which one or more of the distractors could be judged correct and plausibly argued in the final analysis constituted the most frequent type of problem students had to contend with. In many of the examples discussed earlier in this chapter in connection with processing strategies and schema activation, low achieving readers generally and high achieving readers in some measure consistently demonstrated difficulty with questions labeled in this category. In many cases these questions were initially answered incorrectly and in others students could be made to see an alternative position when it was drawn to their attention. On the question from the Ruth passage cited earlier in this chapter a student was shown to have chosen the correct answer for the wrong reason. Notice that in the following interview with another low achieving 3rd grader (AC) essentially the same type of analysis leads instead to the selection of an incorrect answer. To recapitulate, the student was to decide on the question of why the map was made out of metal. The student responded

S: 'So that many people could work on it.'

I: You mean 'on top of it' or what?

S: Yeah, like if you have a big map and you write on it where you're going to go and how you're going to get there.

I: I see. Alright.

I: Does it have to be made out of metal for people to work on it?

S: Not necessarily.

I: What is the question asking you?

S: It's asking why was the map made out of metal. It's metal because if you're driving along and you don't have any mats or anything, you don't have to have a table you can just write on it.

I: Where does it say that they were going to put the metal map? Does it say.

S: I forget.

I: See if you can find it. Where do you think it would be.

S: It didn't say anything about where they were going to put it. (Finally locating the reference). Oh, outdoors, they were going to put it outdoors.

I: Now, if they put it outdoors, do you still think that answer is the right one, so that many people could work on it? See what I mean.

S: Yeah. I see what you mean. People didn't have to work on it, they could just see it, like, "Look at that", we can just see it; we don't have to work on it.

I: But they still could if they had to.

S: But they could if they had to.

I: Any other answers there that make sense to you at all, that could be possible answers?

S: It could last long (sounding skeptical).

In one or two cases students even resisted the expected answer on quite logical grounds. For instance, on the Balloon passage question cited in Chapter 2 exemplifying this category (p. 28), a low achieving 5th grade student (15B) rejected a suggestion that the first balloons were piloted for sport, initially saying

S: There's not too much on that part.

The interview continued

I: When they talked about J.A.C. Charles and Letitia Sage going up in balloons, why do you think they went up?

S: Just to see if it would work, since it was new.

I: Is that the same as any of these over here.

S: 'C'? (that is, instrument testing)? (Sounding doubtful.)

I: Or doesn't it quite fit?

S: Not really.

The logic of her answer is quite defensible from the context provided, but it seems to preclude the possibility of choosing the expected answer.

Difficult inferences. This problem type was identified much less frequently than the preceding type and, overall, constituted less of a difficulty. We have included it as a verified problem type, however, because even though students were often able to find their way to the correct choice, this generally happened by default--a consequence of the question having clearly inappropriate distractors. It was mainly those students who did not use a process of elimination strategy in selecting their choice who answered incorrectly. When a particular question under this category lent itself to the possibility of selecting and justifying one of the distractors, as did, for instance, the question from the Maria passage exemplifying this type (p. 29), this category becomes effectively indistinguishable from the category of plausible distractors. The main difference is that the particular cause of the student's confusion can be more precisely attributed to a structural inadequacy in the passage. This is made clear in an interview with a low achieving 8th grader (CK) on this question. The student initially answered the question incorrectly.

S: 'B' probably, 'complete their calculations of the comet.'

I: Do they actually tell you in the story?

S: No, just that they sent him stuff, I think....It doesn't say what she sent it for.

I: How do you think they expect you to find out? Do they give you any clues?

S: She just sent it and he sent back good news.

I: Is that a clue?

S: Yeah, probably be 'd' (that is, the correct answer).

The student was able to be guided to the correct answer indicating that the necessary information was there in the passage, but that it was somewhat inaccessible because of placement and structuring. In other words the information was apparently not highlighted enough for the student to see the possibility of an inferential link on the first pass.

Unstated point of view. Examples of student difficulties with questions under this type were presented earlier in connection with the discussion of test-wiseness.

Inadequate identification of referents. This problem type was verified as a true difficulty mainly for younger students. Even by the 5th grade students appeared able to skirt the inconsistencies entailed in treating an inadequately specified referent as definite. Third graders, however, were often misled by the lack of specification into assuming the question pertained to a passage character other than the one intended. For instance, on the Ruth passage question illustrating this type (p. 32) two 3rd grade low achieving students (DF, AC) had the same reaction to the expectation implicit in the question that the character under consideration was specifically identified in the passage. Both assumed "the girl" was Ruth Asawa, when in fact she was an unspecified friend of Ruth's. DF initially chose an incorrect answer, but on lookback she changed her answer to the one expected. Even so, however, her understanding was insecure.

I: Do you remember the girl they're talking about? Who is it? Do they tell you?

S: I don't know. Oh yeah, the artist Ruth, she took a wooden knife.

I: The girl they're talking about is Ruth then?

S: There's also another girl and a boy.

I: Is this girl the same as Ruth?

S: I don't know.

I: What do you think about the girl they're talking about here. Is that Ruth?

S: Yeah.

Absence of superordinate topic. Questions in this category occurred relatively infrequently, although they reliably led to

difficulties in interpreting the force of a question. For instance, in the Cosby passage exemplifying this type (p. 34), the absence of an explicit topic statement on which to hang the information led students to interpret the question as asking for a statement of fact rather than a statement of purpose. The following interview with a low achieving 3rd grader (EB) illustrates this point.

S: I don't get that question, 'The story says that Bill Cosby has been on TV shows to....'

I: What do you think it's asking?

S: I don't know. Something like what's he do on TV shows.

I: If that's what you think it means which one do you think is the right answer? Which are you thinking about?

S: I'm thinking about all of them - play football, I'm not sure of; making records...well, yeah, he sings some of his songs on TV shows.

I: So you think that might be the answer?

S: Yeah.

I: If the question here means something a little bit different - if it means 'Why does Bill Cosby go on TV'

S: I guess because the studio likes him.

The strength of the student's initial interpretation of the question evidently makes it difficult for him to pick out the correct answer even when specifically cued.

Unknown vocabulary in question. This category was initially not anticipated, but, even if it had been, there would have been no way to reliably predict which questions contained critical vocabulary unknown to the young readers. Evidently, however, the controls the test developers use to assure that vocabulary which cannot be inferred from the passage context is eliminated sometimes breaks down.

When the problem revealed itself among the answer options students, particularly those among the high achievers, were not inclined to choose items whose meaning they did not know. Presumably, they assumed that, if a vocabulary item was unfamiliar it was one of the distractors. They then set about trying to verify another option as the correct answer. An interview with a high achieving 5th grader (AE) on a question from the Plants passage illustrates these points. The student does not know the meaning of the expected answer, the word 'hardy'.

S: 'From the passage one can tell that most insect-eating plants are very...' What does 'hardy' mean?

I: Well, before I answer you, go through the other possibilities and see whether any make sense.

S: O.K., well, one answer could be 'b' for the bugs for the insects, cause the reason they went to the plants is because they smelled something sweet, and it says 'tasty', and then they just catch the insect.

I: Now, 'hardy' is a word that means strong, and able to withstand hard conditions.

S: That would be another possibility.

I: Is one better than the other?

S: Well, I guess I'd go with 'a' (hardy). 'b-tasty,' that's what sort of lured the insects to the plant, because they smelled something swee[redacted]

I: Which one are you goi[redacted].

S: Probably equally good[redacted] can't make a decision.

I: Now, if you didn't know what 'hardy' meant.

S: Then I'd go with 'b'.

On analysis the interviews reveal that students are not capricious in choosing the answers they do to test questions. Furthermore, evidence of serious comprehension difficulties is lacking. We have seen reason to believe instead that low achieving students evidence a lack of metacomprehension and test-taking skills sufficient to successfully find their way through the structural and lexical difficulties which the tests themselves present. Under this interpretation good readers are those who have an ability to process poorly written prose. It remains to be seen in the next chapter whether we can substantiate this claim by demonstrating enhanced performance when passage material is rewritten to improve its structural cohesiveness.

Chapter 4

EXPERIMENTAL RESULTS

The interviews indicated that low proficiency students apparently can access the same comprehension abilities as high achieving students, but that they lack or do not exercise verificational routines to assess their choice in a fuller context. Because materials are sometimes written in vague, ambiguous, or misleading ways, the verificational procedures become as or more important than comprehension routines in selecting the wanted answer. This suggests that the tests are measuring some test-taking skills in addition to comprehension skills. It also suggests that if the various ambiguities, inconsistencies, and inaccuracies are reduced or eliminated the balance would swing more toward the assessment of comprehension skills than the assessment of test-taking skills. To this point the experiment described in Chapter 3 was designed. The expectation was that by rewriting passages from standardized reading comprehension tests we could show enhanced performance for low achieving students.

Test Objectives

Several considerations had to be met in designing the experimental test to assure that it did not radically change the character of the original test. It was necessary first to assure that the objective of each test question was held constant; for example, that a question calling for an inference remained an inferential question for the rewritten passage as well. It was, of course, difficult to control for subtle differences in the obviousness of the inference, since in many cases the reason for an ambiguity could be traced to the inadequacy of the discourse in providing sufficient evidence of a discourse link. For example, in the Sequoya passage (CAT15C) a question asks why it was harder for the Cherokees to talk together after they had moved to the West. Part of the information cueing the correct answer--that they lived farther apart--must be inferentially linked up. In the relevant part of the original version, the sentences

The Cherokees were pushed from the South to the West.
Their homes were scattered.

leave the temporal link between the moving and the scattering inexplicit, leading to possible interpretation that the two events are simultaneous rather than sequential. In other words the scattering could equally apply to the homes in the South or to the new homes in the West. In the rewritten version, the sequential connection is made explicit.

The Cherokees were pushed from their homes in the South and forced to scatter to the West.

The change removes the need to make this inference, although the inference to the proposition of the Cherokees having difficulty talking to one another remains intact as does the lexical inference between living farther apart and scattering. Consequently, the larger comprehension objective, though modified, is essentially the same.

The desire to keep the question objective intact, however, requires an assurance that the objectives across all passages within the same test are consistently applied and, furthermore, that all of the tests define their objectives in a similar manner. An examination of the tests showed that neither of these assurances is warranted. On the first point it sometimes happens that a question declared to test a certain objective actually meets the criteria associated with another objective. For instance, in the MAT (Elementary JS) it is claimed that the question

Maple syrup is made from tree:
a. bark
b. branches
c. sap
d. leaves

tests literal comprehension of specific passage details, based on the reader's use of "information stated explicitly in the passage" (Prescott et al. 1978:55). However, the information necessary to give the expected answer--sap--is presumably contained in the sentence

Maple sugar comes from the sap that runs through Maple trees.

which makes no mention of maple syrup. The reference to syrup is contained in a separate sentence following a discussion of how sap is boiled to produce it. The reader must make an inference drawing together information in different parts of the passage in order to correctly answer the question. The true objective of this question, then, is closer to what the MAT defines as inferential comprehension of global passage information, namely the use of information "implied in all or a large section of the passage to infer answers to questions related to...cause and effect" (Prescott et al. 1978:56).

On the second point the three tests examined cannot be said to utilize the same objectives, and although superficially there seems to be at least some agreement on the broad objectives, it is apparent that the definitions of these differ. Figure 4-1 lays out the main objectives of the MAT, CAT, and ITBS, giving for each an encapsulated definition and including under some of them one or more examples of narrower objectives. The first point to consider is the general agreement that there are three modes of comprehension based roughly on the nature and scope of material the reader must process in order to answer a question. Only the MAT, which lists six primary objectives, falls outside this pattern, but four of the six can be unified on the two parameters of literal comprehension and inferential comprehension. Doing this reduces the total to four primary objectives, three of which align closely with the objectives on the other tests. We will consider the fourth, vocabulary comprehension, at a later point. The terminology varies somewhat between tests--for instance, the MAT's inferential comprehension corresponds to the CAT's interpretive comprehension and the MAT's evaluation corresponds to the CAT's critical comprehension and the ITBS's generalizations-but we make the equations because each test sets up the same number of distinctions and, as the category breakdown shows, each includes some common sub-objectives.

Closer analysis, however, shows that the main objectives are neither explicitly nor implicitly defined in the same way. The explicit

Figure 4-1

Test Objectives

MAT	
1. vocabulary comprehension	word meaning from context
2. literal comprehension of specific passage details	explicitly stated information
3. literal comprehension of global passage details main idea cause and effect	information based on all or large part of passage
4. inferential comprehension of specific passage details	implied information
5. inferential comprehension of global passage details main idea	information based on all or large part of passage
6. evaluation real and imaginary author attitude propaganda	reader judgments relating to mood and criticism
CAT	
1. literal comprehension recall of facts	actually stated information
2. interpretive comprehension main idea and cause and effect solutions to problems inferred meaning figurative language (i.e. literary allusions) etc.	implied information
3. Critical comprehension real and imaginary persuasive techniques author attitude etc.	reader judgments
ITBS	
1. Facts (literal meaning) vocabulary in context	stated factual details and relationships (WH question)
2. Inferences (interpretive meaning) cause and effect characterization	Why questions
3. Generalizations (evaluative meaning) main idea author viewpoint figurative meaning etc.	higher order skills

definitions are first of all weak. They do not inform the critic about the cognitive scope of an objective. We don't know, for instance, how to interpret the MAT's definition of evaluation--as a reader judgment relating to mood and criticism--in terms of what the reader must actually do to solve the problem. Specifically what does he or she have to do in rendering an evaluation that is different from what happens when a literal or inferential comprehension question is answered? The examples which are included under each category objective, such as an ability to distinguish between real and imaginary events or to perceive the author's purpose, provide implicit definitions of the objective by constraining its range of applications. But even though the examples provide more information as to the meaning of the objective, a comparison of the three tests at this level shows a lack of agreement on where one or another of these subobjectives should be placed. Consider, for instance, the reader's ability to grasp a main idea as a subobjective. In the CAT this ability is considered an instance of interpretive comprehension, in the ITBS a generalization or evaluation, and in the MAT an instance of either literal or inferential comprehension. There is certainly no uniformity of opinion on what comprehension skills underlie this objective. Other examples include the assessment of figurative language, which is considered an interpretive skill by the CAT and an evaluative skill by the ITBS and recognition of vocabulary from context, a literal skill by the ITBS, an independent skill by the MAT, and a skill which goes untested in the CAT.

Obviously the agreement on the components of the main comprehension skills is in some doubt, a consequence perhaps of our poor understanding of the entire comprehension process. If this is the case, the individual tests should not be faulted on their respective interpretation of what makes up a superordinate comprehension skill, though each of them can be legitimately criticized for not applying their own criteria consistently. For the classroom teacher attempting to apply test results to instructional decisions, the best that can be said is that the test objectives may only weakly correlate with real abilities or real deficiencies.

For the experimental procedure the lack of agreement on test objectives and the lack of consistency with which the criteria were applied meant that the point of each question had to be reassessed from a more objectively derived standpoint. To do this we adopted a set of structural criteria based on the extent of the information which had to be appraised before the correct answer could be apprehended. At the same time we wanted to stay within the realm of the categories as they were loosely laid out in the tests.

The resulting analysis set up and defined four comprehension objectives. The first of these, literal comprehension, involves the assessment of information in a single sentence of text; for example an MAT (Primary 2 JS) item asks the reader to complete the sentence

> Fresh mud was used to
> a. bake the bricks
> b. hold the bricks together
> c. make boxes for bricks
> d. put on the floor

The answer is to be found in the following sentence

> The bricks were then taken from the boxes and stuck together with fresh mud.

For practical purposes, a sentence which contained all the information pertinent to answering a question except that it used a pronominal marker in place of an identified referent was still considered to be literally comprehended. Even though an inference has to be made to identify the referent, it was felt that this type of inference is of low order and hence not critically affecting the designation. One of many instances of this type occurs in the following example from the CAT (13D). The question asks

> According to the story where did Bill Cosby play football?

based on the passage sentence

> He tells stories about playing football in the street and racing go-carts.

The he in the passage refers to Bill Cosby.

Inferential comprehension involves the recognition and connection of information contained in two or more sentences of the texts; for example in an item from the Lawn passage (ITBS 7) a question asks

> What makes some people think the lawn is soaked when pudding occurs?

The passage information states:

> Water either forms in the puddles on the surface or just runs off. This is called "puddling." When puddling occurs many people think the lawn is soaked.

Instances of this type are quite common and involve a host of purposes, although all of these are possibly unifiable around the notion that they lead to an appreciation of the logical relations connecting individual sentences to the organization of the discourse.

The third type, evaluative comprehension, requires the reader to appraise passage information in connection with background information, that is, information not available explicitly or inferentially in the passage itself. Questions in this category can range from requests for interpretations, such as judgments of affect, to information which is somewhat tangential to the text. An example of the first type from the Sequoya passage (CAT 15C) asks

> Which of the following best describes Sequoya.

Relevant passage information includes the following scattered sentences.

> Sequoya, a talented Cherokee teacher, was also deeply concerned....Sequoya thought that if he could make a Cherokee alphabet, his people would be able to write to each other. He worked hard for twelve years....The Council decided that Sequoya had indeed learned the secret of the "talking leaves." Within a few months Sequoya was able to teach thousands of his people to read and write their own language.

The correct answer--intelligent--is nowhere explicitly mentioned or alluded to, so that the reader must make the judgment, on the basis of

his or her own background knowledge, that Sequoya's intentions and behaviors were in the class of intelligent acts.

The second type of evaluative question is generally regarded as a literal type by the test makers who use it. It does, in fact, ask for factual information, but the information must be brought up from background experience. An example from the Australia passage (ITBS 8) asks

> The body covering of the greater glider is most like which of these?
>
> a. A cat
> b. A bird
> c. A fish
> d. A snake

The passage provides only the following information.

> Zoologists describe it as a "hairy mammal," which means it has fur...

The question, consequently, requires the student to reflect on his or her background knowledge of animals and, by the convention adopted here, becomes evaluative.

The fourth and last category objective, main idea comprehension, departs from the others in not applying a strictly structural criterion. The notion of main idea, as discussed in Chapter 2, is primarily a pedagogical creation, rather than a discourse reality. In a strict sense any idea can be the main one, depending on what interests and background the reader brings into the reading situation. A test-taker who has not been taught that a main idea is to be found in the various paragraph topics will be at some important disadvantage. For this reason the main idea questions in the tests were put into a separate analytic category, even though they might otherwise have best been assimilated into the evaluative comprehension category.

The resulting scheme of the reconceptualized test objectives is summarized in Figure 4.2. This scheme was used to reassess the objective of each question used in the experimental test, and forms part of the analytic framework for discussing the experimental results. In numerous cases this resulted in the question being recategorized, most often in the direction of supposedly literal comprehension questions being reanalyzed

into other categories. A full documentation of the changes made in test question objectives is included in Figure 4.5.

Figure 4-2: Reformulated Test Question Objectives

literal comprehension	substance of question is retrievable from a single test sentence
inferential comprehension	question is answerable by drawing together text information in two or more sentences
evaluative comprehension	question requires coordination of text and background information
main idea comprehension	question focuses on one or more paragraph topics.

Test Questions

Besides leaving question objectives intact, a second consideration in rewriting passages was to leave the test questions themselves in their original form, even when these were judged to be poorly written or confusing. In such cases, special consideration was given to clarifying the question's intent by rewriting relevant parts of the reading passage, if this was possible without introducing new information or specifically cueing the expected answer. In some instances, however, no intercession could meliorate the situation. For example, questions whose expected answer is an importation from background information cannot be improved. The following question from the Sailors passage (ITBS8) is an example.

Which food would best prevent scurvy?

a. beef
b. grapefruit
c. doughnuts
d. candy

The expected answer—grapefruit--is not mentioned in the passage and is cued only by the sentences

The limes, like most fresh fruit and vegetables, contained Vitamin C. This vitamin both cures and prevents scurvy.

Consequently, without actually making reference to grapefruits in the rewrite and violating the stricture on importing new information into the passage, the answerability of the question could not be improved.

A second type of impervious question was one in which the expected answer contains a critical vocabulary item which is unknown to the student and which is not easily inferrable from the passage. The following question from the Plants passage (CAT 15D) illustrates this type.

From the passage one can tell that most insect-eating plants are very

a. hardy
b. tasty
c. useless
d. colorful

This question, which was also used in the interview sessions, proved difficult for several students because the expected answer--hardy--was not in their active vocabularies, and the passage cues were insufficient for them to infer that this was the expected answer. Instead, these students tended to select the answer option--tasty--justifying their choice with the following passage sentence.

The Venus's fly trap has special juices on its leaves. Insects land on its leaves to taste the juices.

The assumption they make, and there seems no way to preclude its logical possibility, is that the plants are tasty to insects. Again, there was no way to rewrite the passage without explicitly cueing the expected answer, a tactic which would have changed the question objective.

Changes in the passage which were addressed to particular questions were made in the interests of clarifying critical points of information. The objective was not to modify the question in its intent but to make it structurally consistent with the information presentation in the passage. For example, a test question from the Ruth passage (CAT 13C) asks

What did the girl use to get paste out of the bowl?

expecting the student to utilize the passage information

> With a wooden knife, a girl took some paste from the bowl and began to make a copy of her school.

The difficulty with the question lies in its presumption--indicated by the definite article the--that some specifically identified girl is being referred to. A few students during the interviews took this to mean that the question referred to Ruth in some way, since she was the only specifically identified "girl". In the rewrite an attempt was made to smooth the inconsistency by making the person of the girl in question more definite, without incorporating new information. This was done by introducing a sentence with topic force to prepare the reader for the introduction of a referent and then tying the referent to the topic by a type of definite marker. These considerations resulted in the following changes:

> Everyone helped. With a wooden knife one of the girls took some paste out of the bowl and began to make a copy of her school.

The previous example can also serve to illustrate other criteria for rewriting in addition to addressing particular questions. For one, by introducing an explicit topic, it amends a lapse in the coherence of the original passage, the third paragraph of which began with the cited sentence. This paragraph essentially recounted the girl's activities as well as the activities of an unidentified boy. These events, though ultimately integrated, have to be looked on as elements of story detail subsidiary to an unstated topic.

A second purpose of rewriting, to increase the cohesion of passage sentences, is illustrated in the present example by the logical inclusion of the stated referent within the scope of a more inclusive entity. The mention of everyone in the introduced topic is synonymous with a previously mentioned reference to Ruth's friends. The girl and the boy, who are subsequently introduced, are themselves understood as members of the larger class of friends. The original version did not ignore this relationship, but it left it to be inferred entirely from a connection in the preceding paragraph.

The Rewrites

Although some test questions were specifically addressed in the rewrites, for the most part changes were made without regard to the information being tested. The larger consideration was to construct a cohesive and coherent account of the basic passage information. We hypothesized initially that a passage rewritten to eliminate inconsistencies and to clarify the progression of information would enhance readability, make the task of retrieving specific information easier, and ultimately make test questions easier to answer, whether or not they had been targeted for attention in the rewrites. (In a few instances the adoption of this priority, in fact, led to an unintended complexity for the student on particular questions (see below).) The passages, consequently, were regarded as whole and entire elements and whether or not a specific question was affected by rewriting was in part a factor of circumstance. Many questions, in fact, were judged not to have been affected by the rewriting and, in the later analysis performance on these questions was computed separately from that on other questions where the rewriting had made a discernible difference.

As mentioned earlier, the type of alterations made in rewriting passages did not fundamentally change the character of the original versions. They did, however, have an effect in changing the character of the syntax. In general, the changes were not sensitive to the usual readability considerations of keeping sentences short and uncomplicated, especially for younger children. In many cases, where it was judged the sentences were collapsed into a single, usually longer version. A check on what effect this had on computed readability (using the Fry formula) showed that 7 of 14 rewritten passages increased their scores by more than one full grade level. Other passages were not appreciably affected (Figure 4-3). These findings were within our expectations because the increased length in some passage sentences directly enters into the Fry readability equation. Overall, however, the sentences of the rewritten passages showed a greater variability in length than did the sentences in the original version.

Figure 4.3

Readability levels for original and rewritten passages using the Fry formula

	Readability level	
3rd Grade	Original	Rewrite
Ruth	7.1	6.8
Bricks	5.7	5.3
Cosby	5.6	5.7
Ellen	3.9	4.2
5th Grade		
Plants	4.3	6.0
Sailors	6.0	7.1
Andrew	4.5	5.9
Lawn	4.8	5.8
Sequoya	9.1	9.5
8th Grade		
Pyramids	9.4	13.6
Maria	16.9	16.8
Forces	12.0	13.2
Australia	10.0	11.4
Minerals	8.6	8.6

This was not a planned goal, but it emerged naturally from attending to other considerations.

Types of Changes Made

Most of the changes made fell into five main types:

(1) reordering of passage information
(2) restatement of passage information
(3) specification of inferrable information
(4) clarification of paragraph topic
(5) attribution of information

Reorderings produced the most dramatic changes, in that the flow of information could be substantially affected. Most instances of this type of change were made in order to restructure the scope of a topic, either in its relation to other topics within the passage or in its relation to the detail in the passage. The first subtype is illustrated by the Lawn passage (ITBS 7) which begins

> In order to make it grow well, a lawn should be watered the right way. A strong hose spray is bad for a lawn.

These two sentences in the full context of the passage are both topics; the full detail includes a discussion of good watering practices and poor watering practices. Consequently, the two sentences were separated as the heads of different paragraphs and the details appropriately allocated. The first paragraph began

> In order to make a lawn grow well, it should be watered correctly. The best way is to spray or sprinkle the lawn.

and the second

> A strong hose spray is bad for a lawn. The force of the water packs the soil particles together....

The second type of reordering amounts to a recategorization of passage information, motivated by a determination that certain pieces of information are ordered in the original in a way which masks their true discourse relationships. In the Pyramids passage (CAT 18D), for

instance, the concluding sentence of the passage suggests that in that position it is acting as the author's assessment. In truth, it may serve as such, but more importantly it simply recapitulates the essence of the discussion and so has more weight as a topic than as an assessment. Consequently, this sentence, with minor alterations, was moved to the head of the second paragraph where it organizes detail describing some physical characteristics of the Great Pyramid. It begins

> It has been determined beyond reasonable doubt that the Great Pyramid was built more than 4000 years ago and that its builders were advanced in science, mathematics, and architecture. They constructed its four triangular sides to a height of a 40 story building....
>
> Continued research and study may offer a better picture of what the Great Pyramid is, but it still does not tell us why or by whom it was built.

Restatement of passage information is often done in conjunction with other changes in order to adapt the altered information to new locations or new functions. Other times it is done to better adapt a point to its contextual environment. For instance, in the Plants passage (CAT 15D) a discussion of how two varieties of insect-eating plants capture insects by moving their leaves is followed by a paragraph describing how a third variety accomplishes this by another mechanism. In the original the paragraph begins by citing a fragmented list of facts whose relationship to one another is problematic.

> The pitcher plant eats insects, too. It cannot move its leaves the way the other insect-eating plants do. The leaves of the pitcher plant make a sweet smelling juice. Insects come close to the leaves to taste the juice.

In the rewrite, the relationships are made more apparent by restating the information, in the process collapsing the first four sentences into two and introducing cohesion elements which bind the information.

The pitcher plant also captures insects, but not by moving its leaves. It makes a sweet smelling juice which draws insects to it....

In some cases a restatement is made to eliminate an ambiguity or an imprecision. In the Ellen passage (ITBS 7), for instance, the second paragraph has a teacher discovering why Ellen couldn't spell well.

> At last the teacher found out why. She could not see the chalkboard well from her seat.

The she beginning the second sentence is potentially ambiguous in its reference. It logically would seem to refer to Ellen, but coming as it does with a mention of the teacher intervening, it could structurally be identified as the teacher, rather than Ellen. This interpretation would preserve the referent focus, something which the use of the pronoun would suggest anyway. In the rewritten version the ambiguity is eliminated by eliminating the pronouns and restating the intended referent.

> At last the teacher found that Ellen could not read the chalkboard from her seat.

The third type of change specifies information which was left to be inferred in the original version. Since a requirement to draw an inference is in itself not unreasonable and is, in fact, a necessary component of reading comprehension, only those structures which were felt to be imprecise or unnecessarily vague were singled out for rewriting. A particular example is included in the Minerals passage (ITBS 8) in a section listing a number of ways to maintain stores of minerals. The third point is stated as follows.

> Three; related to number two, is to find cheaper methods of mining those deep-lying minerals and of processing ores low in minerals.

The second point, mentioned parenthetically, is to explore for deeper mineral sources. Consequently, the nature of the imputed relationship between two and three is causal. Rather than leave this inference as unnecessarily vague as it is, the sentence was rewritten to read

Three, which is necessary if two is going to be successful, is to find cheaper methods....

The fourth type of change, clarification of the paragraph topic, is related to the preceding type. Both involve a greater specification of presupposed information, but topic clarification has more important functional consequences, since the inference involved can be critically important in schema formation. The Bricks passage (MAT Primary 2 JS) illustrates this point. The original passage, in a single paragraph, discusses something of the historical manufacture and use of bricks and concludes with a reference to the modern use of bricks.

> Although they are now made by machine, bricks continue to be important in the building business.

In the middle of this paragraph, the information abruptly changes in scope from a procedural account of brick making to giving three instances of the usefulness of bricks. The three uses are not linked to any explicit, encompassing statement in the passage, leaving them with only an inferred topic. In the rewrite, the hiatus between the two topical foci was made explicit by putting each in a separate paragraph and making the topic of the second paragraph explicit. The rewrite in part read

> Brick houses had many good points. If they were built on a strong base, they could stand for hundreds of years.

where the first sentence is the inferenced topic made explicit.

In this same passage the concluding sentence also implicitly recognized a separate topic in its mention of machine-made bricks. This topic was also reified and set up to head a separate paragraph.

> Bricks are still made today, although now they are made by machine. They continue to be important in the building business.

This paragraph even as it stands in the rewrite is not completely coherent since it lacks the points of detail contrasting machine-making with hand-making which would make a tie to the final assessment more motivated.

The last type of change, attribution of information, will perhaps be more controversial than the others because, by its intention to clarify the source of information, it may arguably change the author's intent to leave the attribution vague. This type of change was for the most part made in passages with a persuasive purpose, where the reader was asked to assess the validity or factivity of some claim. This entire genre of questions, however, leads to difficulties on the reader's part because questions may be answered differently as one or another point of view is adopted. The aim in rewriting the passage, then, is to eliminate some of the ambiguity by making the point of view clearer. For example, in the Andrew Warrell letter (CAT 15C) the statement

> It's the best show Oliver School has ever had.

leaves it unclear if this is the letter writer's personal opinion or his reporting of the student body's consensus. To clarify the situation, the sentence was rewritten to include the pronoun we referencing the author and the club he is noted as representing.

> We think that it's the best show Oliver School has ever had....

The reader must still infer the identity of the we, but the likelihood of it referring to the student body is lessened.

Effect of Rewrite on Questions

After all passages had been rewritten a determination was made for each question as to whether the rewrite had affected the answerability of a question in any way. Underlying this procedure was a desire to distinguish the composite results from the set of results on the affected questions. The larger purpose was to provide some gross indication of how the cumulative affect of rewriting a passage differed from the question-specific effects, whether, for instance, the rewrite as a whole made it easier for the student to locate information relevant to answering even unaffected questions.

The problem was complicated at the outset by the difficulty in deciding whether a particular passage change had indeed altered the

probability that a question might be answered correctly. Some changes which had been made for reasons other than their impact on a particular question could nevertheless exert a subsidiary effect on that question. For instance, a reordering of information in the Bill Cosby passage (CAT 13D) for the purpose of clarifying the topic-detail relationship possibly altered the interpretability of a test question. The question read

> Bill Cosby's stories show that he probably was
>
> a. happy about being young
> b. excited about watching TV
> c. worried about his brothers
> d. curious about making records

with the expected answer the first option. The original version cued the answer with the following sentences, comprising the last paragraph of the passage.

> Bill Cosby likes children and they like him. His stories help everyone remember what it was like to be young.

The rewrite puts Cosby's stories, the passage theme, into topic prominence as follows.

> Bill Cosby's stories show that he likes children and the things they do. They help everyone remember what it was like to be young.

The required evaluative demands of the question have not changed, but the match in wording between the question and the rewritten passage is closer, perhaps making it easier for a student to draw the necessary inferences. Because of this possibility this question was considered to be affected by the rewrite, even though only minimally.

Judgments on the effect of a rewrite were subtle enough in some instances that two raters were used to independently assess the effect on each question. Initial differences in judgment, which occurred on approximately 15 percent of the questions, were resolved in conference to the satisfaction of both raters.

In a few cases it happened that the rewrite had a negative effect on the answerability of a question, though this went undetected until after the experiment had been run. For questions so affected the rewrite had inadvertently made one of the distractors a logically possible answer, in addition to and in competition with the expected answer. In the Ruth Asawa passage (CAT 13C), for instance, a change in the second paragraph for the purpose of making the topic more precise ended up highlighting one of the distractors. The original version began the paragraph as follows.

> First she drew a map on paper. Then she and her friends began working.

Since the drawing of the map itself constitutes work, it should fall within the scope of the second sentence. The rewrite then attempts to restate the scope of the work by more clearly specifying the sequence of events. It took the following form.

> She called her friends together and they began working. First, Ruth Asawa drew the map on paper.

One of the questions on the passage focused on the sequencing of these activities. It asked

> In this story, which of the following did Ruth Asawa do first?
>
> a. made the paste
> b. made a metal map
> c. talked to the children
> d. drew the map on paper

The expected answer, the fourth alternative, is still overtly cued in the same way as in the original passage, namely by the word first. However, the rewrite also makes the third option a possibility by stating that she called her friends. Since this event precedes the map drawing, students could ignore the overt, literal cue and interpret this as an inferential question instead. These students were, in fact, being unfairly penalized for their greater perceptivity.

A total of 9 questions (out of 85) were judged to be negatively affected by the rewrite. Because at one stage of the analysis these questions were eliminated from consideration, the rationale for the decision on each of them is provided below. The first question, from the Ruth passage was discussed above.

2. Plants Ques. 60: What happens when insects land on the sundew?

   Expected answer: They stick to its droplets.
   Highlighted distractor: They drown in its juices.
   Rewrite: When an insect lands and gets stuck on one of the drops, the sundew's "arms" close around it.
   Comment: This question was judged to have no best answer since the reader would expect that the end point of the process would be the targeted answer. This was not included among the options. The rewrite rather than clarifying the situation aggravates it by implying that the process of getting stuck is presupposed background information rather than the looked-for result. The highlighted distractor is an inferrable consequence of the insect landing on a drop of fluid.

3. Sailors Ques. 38: Why did the sailors become tired and weak?

   Expected answer: They did not eat the right food.
   Highlighted distractor: They did not have enough to eat.
   Rewrite: They would become so tired and weak they could not work. The men in charge did not know what could be wrong. They ruled out food because the sailors got plenty of meat and bread to eat.
   Comment: The last sentence quoted above leaves open the possibility under one interpretation that the men in charge eliminated (ruled out) food, implying that the sailors didn't have enough of the right food to eat.

4. Andrew Ques. 44: Which of the following is a fact about the "special prize"?

Expected answer: Andrew Warrel wants the club to receive the prize.

Highlighted distractor: Oliver School won the prize for their decorated halls.

Rewrite: In fact, we're hoping that the Pen and Ink Club will be given the school's special prize.

Comment: Because the prize was definitized in the rewrite, some readers evidently came to interpret it as something already in existence rather than as a hypothetical prize. The highlighted distractor is then the only plausible option which could constitute a fact about the prize.

5. Lawn. Ques. 60: Which would be the best name for this article?

Expected answer: "How to water a lawn"

Highlighted distractor: "A growing lawn"

Rewrite: In order to make a lawn grow well, it should be watered correctly.

Comment: Evidently the juxtaposition of lawn and grow in the first sentence of the passage cued the distractor. In the original version the first sentence juxtaposed lawn and watered.

6. Sequoya Ques. 66: When the Cherokees moved to the West, it was harder for them to talk together because

Expected answer: they lived farther apart.

Highlighted distractor: they had to learn a new language.

Rewrite: The Cherokees were pushed from their homes in the South and forced to scatter to the West... Sequoya knew it would be hard for families and friends to communicate with each other. The Cherokee language had been spoken for centuries, but the people did not have a way to write it.

Comment: In the original passage the information is given that the homes of the Cherokees were scattered probably making it easier to infer the expected answer. In the absence of this clue in the rewrite, some readers evidently chose to interpret the information that the Cherokees learned to write as implying that they had to learn a new language.

7. Maria Ques. 34: Before Maria Mitchell could claim to have discovered a comet, she had to

Expected answer: make a series of observations.
Highlighted distractor: calculate how far the comet was from the star Polaris.
Rewrite: ...Maria noticed an unusual, fuzzy white patch near the star Polaris. She suspected that this might be a major discovery, and for several nights she and her father observed the patch as it became clearer.
Comment: The rewrite reorders events so that the usual observations were followed by a suspicion of discovery. The original, however, removes the element of suspense and states immediately, "...while making her usual observations Maria made a major discovery." In the rewrite, then, the association between the proof-positive discovery and the observations is not as apparent and readers were freed to infer a reason in the highlighted distractor.

8. Australia Ques. 88: What is the main purpose of the second paragraph?

Expected Answer: To tell how the glider flies.
Highlighted distractor: To compare the glider with a bird.
Rewrite: It does not flap its membrane like a bird would flap its wings...
Comment: In filling out the details of an analogy which the original leaves only loosely specified, the rewrite seems to make an implicit comparison more consequential.

9. Australia Ques. 89: Which of these phrases is closest to the meaning of the word "launch" as used in the second paragraph?

Expected answer: To shove or send off.
Highlighted distractor: To set afloat.
Rewrite: ...it launches itself from a high tree and sails.
Comment: The original version specified that the glider sails through the air, a point which being absent in the rewrite sets up the possibility of floating rather than a directional motion.

All in all the probability that the rewritten passages disposed readers to other than the expected answers points up if nothing else the difficulty in designing questions to be unambiguous. It offers a lesson in humility, and certainly must put one in greater sympathy with the test developers. Fortunately, however, the examples can be used to test the interesting hypothesis that readers are in fact sensitive to such subtle changes in structure. Specifically, we would expect that the response pattern on the items would differ between the original and rewritten versions, with the rewritten versions showing the effect of highlighting one of the distractors over the other options. In the original versions there would presumably be not as much reason other than chance for the highlighted distractor to be selected.

As the data tabulated in Figure 4-4 demonstrate , this expectation is met. All of the questions, except the one from the Ruth passage, show a reduction in the percentage of correct responses, with the reduction accounted for mainly in increased response to the highlighted distractor. Furthermore, the pattern of incorrect responses changes in the predicted direction. In the Ruth passage, for instance, distractor 2 which was favored 40% of the time in the original was not selected at all in the rewrite. On the other hand the highlighted distractor was chosen in the rewrite but not in the original. Altogether 6 of the 9 questions show a similar shift in response pattern, and 2 of the remaining 3 show an increase in the percent response for the highlighted distractor. Only the Lawn passage question shows essentially the same response pattern between the two versions. The data, in short, seem to support the hypothesis of reader sensitivity to structural changes. In the following section we will present the remainder of the experimental results with the aim of providing further substantiation for this hypothesis.

Figure 4-4

Questions for which the rewritten passages negatively affect the probability of selecting the expected answer. The item analysis presents the percent of the low achieving students responding to each option. (The Distractor 1 heading for the original passage corresponds to the option listed as Highlighted distractor for the rewritten passage. The distractors listed as 2 and 3 under each category likewise correspond.)

	Original				Rewrite			
		Distractors				Distractors		
	Expected answer	1	2	3	Expected answer	High-lighted	2	3
Ruth, Question 23	60	0	40	0	75	25	0	0
Plants, Question 60	50	17	17	17	50	50	0	0
Sailors, Question 38	100	0	0	0	60	20	20	0
Andrew, Question 44	75	0	25	0	67	33	0	0
Lawn, Question 60	60	40	0	0	40	40	20	0
Sequoya, Question 66	40	40	20	0	0	80	20	0
Maria, Question 34	100	0	0	0	50	50	0	0
Australia, Question 88	40	0	40	20	33	50	17	0
Australia, Question 89	100	0	0	0	50	17	33	0

Analysis of test data

Before beginning the analysis of the experimental test data, it should be remembered that the passages used were purposely selected for their diversity of purpose and type. This had the effect of making the experimental test conform more closely to a typical reading achievement test and permitted a variety of test objectives to be examined, but it also had the effect of limiting the comparability of passages. In fact, few individual passages were balanced across all the objectives and some tended to focus almost exclusively on a single objective. Then, too, even though we put no great reliance on the supposed usefulness of readability formulas to assist in making decisions about passage difficulty, certain rewritten passages may possibly introduce structural complexities beyond those present in the original versions. In a sense we may have achieved a reduction in the complexity of the discourse structure of the test passages at the expense of introducing greater syntactic complexity. The effect of this kind of trade off is predicted to benefit comprehension overall, but it could happen that the absolute size of the predicted gain is tempered somewhat by the increased syntactic complexity of the rewritten passages. The experimental design however, was not sensitive enough to decide this matter, although there is some evidence, to be discussed later, that the greater syntactic complexity may in fact have had some depressive effect.

The unresolved question of the effect of increased syntactic complexity coupled with the small number of subjects and the large number of structural variables influenced during rewriting complicate the statistical analysis of the data; making it improbable that gains or losses large enough to demonstrate statistical significance would be achieved. Consequently, the meaning of any particular statistic must be qualitatively assessed by considering the probability of the gain (or loss) in measured comprehension against the potential influence of unmeasurable competing effects. Admittedly, this is not as satisfying an analysis as would be a demonstration of statistical

significance, but at the least it can provide some indication of the likelihood that tighter controls and the inclusion of more subjects would produce the desired effects.

For ease of comparison and analysis the cumulated scores of all subjects responding to particular questions are presented in figures 4-5a,b,c together with information characterizing each question for its declared and actual comprehension objective and the effect the rewrite had on the answerability of the question. In addition, the discourse type of each passage is provided. It will be recalled from earlier discussion that the declared comprehension objective is that which is specified in the respective test manual, while the actual objective is that which was determined by our own analysis. The effect of rewrite has three possible values: positive in the case that a change had a predictably enhancing effect on comprehension, negative, if a predictably depressing effect, and null if the particular question was unaffected.

As the three tables show, the results on individual questions are equivocal, other than that they verify the proficiency criterion used to select subjects. In other words, for all but a very few questions the low achieving group scored lower than the high achieving group on all passages. However, a test of significance for the difference in proportion of low and high achieving students answering correctly indicated that the differences were not statistically significant for certain passages in their original versions (see Figure 4-6). Since we can assume these differences to have been significant in the norming sample, we have some numerical verification of our claim that the subject sample was too small to yield statistically significant data. It will, then, be possible to show significantly enhanced performance on the rewrites only for those passages in which the difference in performance on the original versions was significant to start with.

If the measure of enhanced performance is taken to be a meaningful reduction in the variability between the high and low achieving groups, then we are able to demonstrate such a reduction only in the cases of the Bricks and Andrew passages (see Figure 4-6). On the

Figure 4-5a

Background Data Summary and Results on Experimental Items

3rd Grade

	Discourse Type	Question Category Objective Declared	Actual	Effect of Rewrite	Percent Correct Response Low O	Low R	High O	High R
Ruth Asawa (CAT 13C)	biography & procedure							
23		literal	literal	negative	60	75	100	80
24		literal	literal	positive	80	75	100	100
25		interpretive	main idea	positive	40	75	40	60
26		interpretive	inferential	positive	60	75	80	80
27		interpretive	inferential	positive	40	25	80	100
Bricks (MAT Pr 2 JS)	history & procedure							
36		literal	literal	positive	20	67	100	100
37		literal	literal	positive	40	67	100	80
38		literal	literal	positive	40	100	100	80
39		literal	literal	null	40	67	100	100
40		literal	inferential	positive	40	60	100	80
Bill Cosby (CAT 13D)	biography							
11		literal	inferential	positive	50	100	100	100
12		literal	literal	null	100	67	100	80
13		interpretive	inferential	positive	0	33	100	100
14		interpretive	main idea	positive	75	50	100	100
15		interpretive	evaluative	positive	40	33	100	100
Ellen (ITBS 7)	biography and explanation							
5		literal	inferential	null	71	25	100	80
6		inferential	evaluative	null	71	50	100	100
7		literal	literal	null	86	50	100	100
8		inferential	evaluative	positive	30	25	67	80

Figure 4-5b

Background Data Summary and Results on Experimental Items

5th Grade

	Discourse Type	Question Category Objective Declared	Question Category Objective Actual	Effect of Rewrite	Percent Correct Response Low O	Low R	High O	High R
Plants (CAT 15D)	categorization and description							
58		interpretive	evaluative	positive	67	75	67	75
59		interpretive	inferential	null	100	100	83	100
60		literal	literal	negative	50	50	100	100
61		literal	literal	positive	67	100	100	100
62		literal	literal	positive	67	100	100	100
63		interpretive	main idea	null	83	75	100	100
64		interpretive	inferential	positive	60	50	100	100
65		literal	literal	positive	60	75	83	100
66		interpretive	inferential	positive	80	75	83	100
Sailors (ITBS 8)	history and explanation							
38		inferential	inferential	negative	100	60	83	80
39		literal	inferential	positive	60	80	100	100
40		literal	inferential	positive	60	40	100	80
41		evaluative	evaluative	positive	100	80	100	100
42		literal	inferential	positive	80	80	100	100
43		evaluative	evaluative	positive	80	60	100	100
44		literal	inferential	null	100	80	100	100

Figure 4-5b (Continued)

5th Grade

	Discourse Type	Question Category Objective Declared	Actual	Effect of Rewrite	Percent Correct Response Low 0	Low R	High 0	High R
Andrew (CAT 15C)	persuasion							
44		critical	evaluative	negative	75	67	100	80
45		critical	evaluative	positive	25	50	100	80
46		critical	inferential	positive	75	83	100	80
47		critical	evaluative	positive	25	67	100	100
48		critical	evaluative	positive	0	67	100	80
Lawn (ITBS 7)	procedure							
60		evaluative	main idea	negative	60	40	100	80
61		literal	inferential	null	80	75	80	100
62		literal	inferential	positive	20	40	100	100
63		inferential	inferential	positive	20	60	100	100
Sequoya (CAT 15C)	history and explanation							
58		literal	literal	positive	80	80	100	100
59		literal	literal	null	80	20	80	80
60		interpretive	inferential	positive	0	0	40	100
61		interpretive	inferential	null	80	60	100	100
62		interpretive	inferential	positive	60	60	100	100
63		interpretive	inferential	positive	60		100	100
64		interpretive	evaluative	null	80	80	100	100
65		interpretive	main idea	positive	60	100	100	100
66		interpretive	inferential	negative	40	0	83	100

Figure 4-5c

Background Data Summary and Results on Experimental Items

8th Grade

	Discourse Type	Question Category Objective Declared	Actual	Effect of Rewrite	Percent Correct Response Low 0	Low R	High 0	High R
Pyramids (CAT 18D)	history and explanation							
64		literal	literal	positive	50	40	80	80
67		interpretive	main idea	positive	50	80	100	100
68		interpretive	inferential	null	25	0	80	100
69		interpretive	inferential	positive	75	80	100	100
70		interpretive	inferential	positive	0	20	80	100
Maria Mitchell (CAT 18C)	biography							
31		literal	literal	null	80	75	100	100
32		interpretive	evaluative	positive	60	75	100	80
33		interpretive	evaluative	positive	80	75	100	100
34		interpretive	inferential	negative	100	50	100	80
35		interpretive	inferential	null	40	50	80	80
36		literal	literal	positive	75	100	100	100
37		interpretive	evaluative	positive	40	33	80	100
Forces (CAT 18D)	persuasion							
58		critical	evaluative	null	40	50	100	83
59		critical	evaluative	positive	40	25	100	100
60		critical	evaluative	null	40	0	100	100

Figure 4-5c (Continued)

8th Grade

	Discourse Type	Question Category Objective Declared	Actual	Effect of Rewrite	Percent Correct Response Low O	Low R	High O	High R
Australia (ITBS 8)	categorization and deacription							
84		inferential	inferential	positive	60	50	100	100
85		literal	inferential	null	60	83	100	100
86		literal	literal	positive	80	83	100	100
87		literal	evaluative	positive	40	33	83	75
88		evaluative	main idea	negative	40	33	100	100
89		literal	inferential	negative	100	50	100	75
90		literal	evaluative	positive	60	67	100	100
91		literal	evaluative	null	20	50	83	100
92		evaluative	main idea	positive	60	50	100	100
Minerals (ITBS 8)	categorization and explanation							
125		evaluative	main idea	null	67	80	40	80
126		evaluative	main idea	positive	67	80	40	80
127		evaluative	inferential	positive	50	60	100	100
128		literal	inferential	null	33	60	100	100
129		evaluative	main idea	positive	40	60	60	60
130		evaluative	evaluative	positive	50	80	100	100
131		inferential	inferential	null	50	20	100	80
132		evaluative	evaluative	positive	17	40	100	100

Figure 4-6

z scores for the difference in proportion of correct response between high and low achieving students on two versions of the tested passages.

3rd Grade	Original		Rewrite	
Ruth	1.7	(n.s.)	1.5	(n.s.)
Bricks	3.6	($p < .01$)	1.3	(n.s.)
Cosby	4.3	($p < .01$)	3.4	($p < .01$)
Ellen	2.4	($p < .05$)	4.2	($p < .01$)
5th Grade				
Plants	1.6	(n.s.)	1.9	(n.s.)
Sailors	1.7	(n.s.)	2.2	($p < .05$)
Andrew	5.6	($p < .01$)	1.3	(n.s.)
Lawn	4.2	($p < .01$)	3.5	($p < .01$)
Sequoya	2.3	($p < .05$)	4.0	($p < .01$)
8th Grade				
Pyramids	3.7	($p < .01$)	4.5	($p < .01$)
Maria	2.3	($p < .05$)	2.1	($p < .05$)
Forces	5.6	($p < .01$)	5.3	($p < .01$)
Australia	3.3	($p < .01$)	3.2	($p < .01$)
Minerals	2.4	($p < .05$)	2.3	($p < .05$)

other hand it appears that using the same criterion the rewrite was detrimental to comprehension in the case of the Ellen, Sailors, and Sequoya passages. The other passages showed no change in the level of their significance between the two versions. These results are reiterated in the summary of data in Figure 4-7a giving the percent correct response across all questions on a particular passage. It would seem; then, that the hypothesis that passages can be rewritten to enhance performance holds in some cases and fails in others.

However, before accepting this conclusion the data should be sifted to focus the analysis on just those questions which were affected by the rewrite. This was done in a two-step process, first eliminating the null questions (Figure 4-7b) and then, subsequently, all the questions judged to have a negative effect on comprehensibility (Fig. 4-7c). Following this procedure we realize two advantages: the number of passages showing a decrement in proficiency for low achievers decreases from six to four and the variance in percent correct response between original and rewritten versions increases for low achievers.

This result indicates that much of the reason for depressed performance on certain rewritten versions is attributable to questions which were either not addressed in rewriting or were negatively affected by the rewrite. What this conclusion further suggests, however, is that the rewrites, may have introduced syntactic or structural complexity of a degree which hindered the student in answering questions which had not been specifically clarified in the rewrite. If this is the case, however, the increased syntactic difficulty does not correlate simply with greater sentence length, since there is no correlation between an increase in readability score (see Figure 4-3) of a rewrite and greater difficulty. Notice, for instance, that the Pyramids and Plants passages, which showed the largest increases in presumed difficulty using the Fry formula are not among the passages showing a decrease in proficiency.

It should be pointed out that the analysis of data for the high achieving students does not pattern the same way as the data for low achieving students. Most importantly, high achieving students showed

Figure 4-7a

Percent correct response: all passage questions

3rd Grade	Low Achieving		High Achieving	
	Original	Rewrite	Original	Rewrite
Ruth (Q=5)	56	65	80	84
Bricks (Q=5)	36	72	100	84
Cosby (Q=5)	53	57	100	96
Ellen (Q=4)	65	38	92	90
5th Grade				
Plants (Q=9)	70	78	91	97
Sailors (Q=7)	83	69	98	94
Andrew (Q=5)	40	67	100	84
Lawn (Q=4)	45	54	95	95
Sequoya (Q=9)	60	53	89	98
8th Grade				
Pyramids (Q=5)	40	44	88	96
Maria (Q=7)	68	65	94	92
Forces (Q=3)	40	33	100	94
Australia (Q=9)	58	55	96	94
Minerals (Q=8)	47	58	80	88

Figure 4-7b

Percent correct response: questions affected by rewrite

	Low Achieving		High Achieving	
3rd Grade	Original	Rewrite	Original	Rewrite
Ruth (Q=5)	56	65	80	84
Bricks (Q=4)	35	74	100	85
Cosby (Q=4)	41	54	100	100
Ellen (Q=1)	30	25	67	80
5th Grade				
Plants (Q=7)	64	75	90	96
Sailors (Q=6)	80	67	97	93
Andrew (Q=5)	40	67	100	84
Lawn (Q=3)	33	47	100	93
Sequoya (Q=6)	50	53	87	100
8th Grade				
Pyramids (Q=4)	44	55	90	95
Maria (Q=5)	71	67	96	92
Forces (Q=1)	50	40	100	100
Australia (Q=7)	63	52	98	93
Minerals (Q=5)	45	60	80	88

Figure 4-7c

Percent correct response: questions affected positively by rewrite

	Low Achieving		High Achieving	
3rd Grade	Original	Rewrite	Original	Rewrite
Ruth (Q=4)	55	63	75	85
Bricks (Q=4)	35	74	100	85
Cosby (Q=4)	41	54	100	100
Ellen (Q=1)	30	25	67	80
5th Grade				
Plants (Q=6)	67	79	89	96
Sailors (Q=5)	76	68	100	96
Andrew (Q=4)	31	67	100	85
Lawn (Q=2)	33	50	100	100
Sequoya (Q=5)	52	63	88	100
8th Grade				
Pyramids (Q=4)	44	55	90	95
Maria (Q=4)	64	71	95	95
Forces (Q=1)	50	40	100	100
Australia (Q=5)	60	57	97	95
Minerals (Q=5)	45	60	80	88

depressed scores on certain rewritten passages wh[illegible] low achieving students demonstrated great improvement on the Bric[illegible] and the Andrew passages, in particular. Where the low achieving students significantly improved their performance on Bricks ($p < .05$) and nearly so on Andrew, the high achieving students lost 15 percentage points on each - a non-significant, but troubling reversal from expectation. The reason for this discrepancy is not entirely clear. It may be a skewing effect of high achieving students performing at or near the ceiling level or it may be attributable to a real difficulty arising out of the rewrite, presumably because the form of the rewrite makes it less obvious that a question falls within a certain category. All except one of the Bricks questions, for instance, require literal comprehension. It may be that the rewrite makes the obviousness of this task in the original version less obvious in the rewritten version. The low achieving students presumably escape the dilemma because the clues to the correct answer as in the original version are not as apparent to them as they are to the high achievers. If this explanation holds it would indicate another difference in comprehension strategies between the two groups, the higher achieving students acting as experts whose attention is directed to points of information which in their oddity of placement stand out more prominently. The low achieving students because they are not expert miss seeing these points of information as out-of-the-ordinary and, in fact, are misled by the poor discourse organization into not seeing them at all. Any improvement in organization, consequently, is likely to benefit their comprehension.

Although most of the analysis was confined to individual passages, a cross-passage analysis of reading objectives using the breakdown arrived at earlier was undertaken. The rationale for this analysis, while not deriving from passage comparability, was assumed to derive from the comparability of questions asked of these passages. Questions purportedly testing similar objectives set up similar expectations in the reader's mind and possibly activate similar search procedures.

Some evidence for this claim was provided in discussion of the interviews.

Figure 4-8 provides the results of data analysis on the experimental test results, taking as an independent variable the actual comprehension objective of each test item as tabled on Figure 4-5. For each objective the number of questions exemplifying it is included in parentheses following its listing.

Figure 4-8 is of interest because it provides some greater understanding of what was accomplished by the rewrites. It shows, first, a difference between grades on the relative proportion of questions in a particular category--almost twice as many literal comprehension questions in the 3rd grade as in the 5th and 8th grades. For the two higher grades proportionately more weight is given to inferential and evaluative questions. This difference is of some consequence because we also find that literal comprehension questions seem to have benefitted substantially more from rewriting than the other types. In fact, at the 3rd and 5th grades improvement in literal comprehension comes very close to reaching significance. This would indicate the positive effects of attention to passage organization to be more apparent in the early grades. Obviously, however, this claim must be verified with larger samples, since the majority of 3rd grade literal questions were drawn from a single passage.

The most disappointing finding revealed in Figure 4-8 is the negligible or even detrimental effect of the rewrites on inferential questions at the 5th and 8th grade levels. Only the low achieving 3rd graders showed improvement on this category. Why this pattern emerged is problematic. It may indicate that the nature of the information organization in the higher grade passages was not as susceptible to change without violating the stricture on changing or importing information. In this event the changes made would necessarily be trivial or inconsequential. Evidence to verify this point, however, is inconclusive, as a judgment regarding the relative effect of a change is somewhat subjective.

Figure 4-8

Percent correct response by question objective: questions affected by rewrite.

	Low Achieving		High Achieving	
3rd Grade	Original	Rewrite	Original	Rewrite
literal (Q=5)	48	77	100	88
inferential (Q=5)	39	58	92	92
main idea (Q=2)	56	60	70	80
evaluative (Q=2)	36	30	70	90
5th Grade				
literal (Q=5)	58	81	96	94
inferential (Q=13)	53	53	94	95
main idea (Q=2)	60	70	100	90
evaluative (Q=7)	56	66	94	88
8th Grade				
literal (Q=3)	69	73	94	93
inferential (Q=6)	58	52	96	96
main idea (Q=5)	52	56	81	87
evaluative (Q=8)	48	57	93	97

In other respects the results do show a rise in performance for evaluative questions at the 5th and 8th grades for low achievers. This is noteworthy because the changes made to clarify many of these questions were relatively uniform. They tended to impart a greater specificity to the point of view the reader was to assume in answering. The results at the 3rd grade are inconclusive. The decrease may simply be attributable to having only two questions in the sample, or it might be due to a real inability of 3rd graders to perform the kind of synthesis of passage and background information required of an evaluative question.

All in all the results of this part of the investigation provide qualified support that rewriting test passages to improve discourse organization can effect an improvement in comprehension for low achieving students. The actual gain in achievement, however, seems to be dependent on the type of passage being manipulated as well as on the type of question asked of the passage. The magnitude of the gain, while significant in only one of the tested passages, must be understood to have been tempered by the small size of the subject sample and, more importantly, by the decision not to change the passage questions asked on the original tests. We demonstrated earlier that many of these questions exhibited faults in their construction over and above whatever faults could be attributed to passage organization. The implication is that by designing better reading passages and better sets of questions, low achieving students would exhibit more evidence of their real grasp on comprehension skills, and less of their lack of test-taking and metacomprehension skills.

Chapter 5

SUMMARY AND CONCLUSIONS

The evidence of the interviews and experimental test is strongly indicative of two major conclusions. First, poor reading ability as determined through standardized test administration has not as much to do with deficient comprehension skills as it does with deficiencies in test-taking and metacomprehension skills. Poor readers of the caliber we looked at in this study were typically able to find their way to plausible but incorrect answers through the application of comprehension skills which were qualitatively no different from those used by better readers. Both types of readers adequately demonstrated their ability to extract explicitly stated factual information, to draw inferences, and to make evaluations of passage content against their background experience. Differences on these skills when they were recognizable were attributable either to specific rather than generic comprehension deficiencies, such as not understanding that the deitic word 'this' can point to a preceding rather than following sentence, or to an overreliance on imported background information to justify answers. Both of these characteristics, however, amount to quantitative rather than qualitative differences between the two populations and would argue for greater attention being paid to discriminating students on their approach to reading and their strategy toward problem solving than on their ability to perform basic cognitive operations on written material.

The second conclusion is in part generated from the argument of the first. If good and poor readers are not so much to be discriminated on the basis of their comprehension skills, then the tests which purport to measure these skills are themselves deficient in not suggesting the real reasons lying behind performance variability. Moreover, the tests are open to criticism in that some of their questions permit two or more logically defensible choices. Since we anticipated the possibility of multiple interpretations from a linguistic analysis of test items and verified these ambiguities through the interviews and experimental tests, we can conclude that the tests, if

their intention is to produce items with only a single defensible choice, have failed in their intention. We have laid the blame for this at two doors: at the test-maker's since their inattention to the linguistic structure of reading passages creates the opportunity for multiple interpretations and at the test-takers' for not understanding how to recognize and resolve ambiguities. It is a real probability, however, that the test-takers' fault is at least partly to be shared by educators who do not themselves recognize children's true difficulties with reading and try instead to remediate problems which do not exist.

Even if it appears unlikely that standardized tests discriminate students on the basis of their comprehension skills, the numbers do indicate that they discriminate students on some variable and, moreover, that they do it reliably. We have already said 'that these true discriminants are, broadly speaking, metacomprehension and test-taking skills. In particular, it would appear that the main difference between good and poor readers involves a reluctance or an inability on the part of poor readers to process information hierarchically. Given that the content of the selections students read on tests is generally unfamiliar, hierarchical processing - a stragegy which organizes and assigns priorities to passage information - is probably more effective in aiding information retrieval than a processing strategy which assigns equal weight to all passage information. The latter what we have called detail processing, is presumably better assumed when a reader already has an elaborated schema on the passage topic. In that event the further details of the passage information find their proper importance through the implicit interaction between the inplace and the incoming information.

Given the particular demands of the reading test situation, it would seem, then, that it is a mistake for students to adopt a detail processing strategy. It may assure that some items are answered correctly, but overall it complicates the retrieval process because information is difficult to locate on lookback. It is possibly for this reason that the poor readers we investigated were generally unwilling to verify their answers. It would have been a time consuming task, possibly not worth the effort in any event.

The etiology of poor readers' preferences for processing written test material is a matter of some speculation. In the minds of some interviewed students, it seems to precipitate from a feeling that tests and reading are somewhat capricious. You can get them to mean whatever you want, and the choice of a correct answer is attributable to the insider's ability to crack the system. They may in fact be right on this point, but that still leaves it problematic how this defeatists attitude arises, especially since they have the necessary cognitive "raw material" to perform adequately.

Perhaps one place to look for the answer is in the tests themselves, which contrary to test-taking strategies which are advocated in instructions and by teachers, actually encourage detail processing by asking for trivial and often times unhighlighted information. Such practices encourage undue attention to these points of information and make them appear important not so much for their intrinsic value to the passage content as for their extrinsic value to the test-makers. Unfortunately, this convention is adopted early on in the child's exposure to tests, perhaps in the mistaken belief that young children are unable to cope with inferential or evaluative questions. Literal questioning becomes the standard for testing in the early grades and in all probability induces the detail processing reading strategy which later so inadequately serves the maturing reader.

In consideration of what can be done to improve the situation, the study investigated what effect rewriting test passages to make them better conform to theoretically determined standards of good discourse structure would have on enhancing performance. As discussed in Chapter 4, the results were somewhat equivocal, although overall the principle was demonstrated, especially if mitigating factors were considered. These included research design decisions, such as not altering the form or substance of certain recognizably poor test questions and not importing information into test passages, as well as evidence that the discourse type of a passage makes it more or less susceptible to improvement, especially in consideration of the type of questions asked of it.

Besides these controllable aspects of the problem, however, there is a set of subject dependent variables, such as motivation, interest, background knowledge, and preferred problem solving strategies, which must also be considered. To the extent that low achieving readers are deficient in one or more of these matters and recognize themselves to be, it is likely that its negative effects on test performance have been carried through on both the original and rewritten versions. It is hoped that the rewritten versions in their greater clarity of presentation may have made students less apprehensive and more sure of their judgments, but evidence of this was not adduced.

Assuming for the sake of discussion that the hypothesis holds by virtue of acknowledging the persistence of these discriminating factors, then the experiment yields some information that improvement is mainly limited to requests for information of a literal or evaluative nature. Inferential knowledge seems less susceptible to the kind of clearly demonstratable improvement as exhibited for the other types. The implication for the testing of reading skills per se is that inferential processing appears less dependent on the stylistic form a passage takes. This conclusion indicates that test developers would more accurately be assessing reading skills if they confined their attention exclusively to inferential questions. It may be noted in this connection that this is what cloze tests, such as the Degrees of Reading Power, already do.

Testers might object that a test ignoring literal and evaluative skills misses the opportunity to inform teachers on important components of their reading curriculum. As well-intentioned as this concern might be, it must be admitted in light of this study's results that the breakdown of objectives on standardized reading comprehension tests has no power to inform on the true status of a student's abilities. Standardized test results cannot be used as the basis for a diagnosis of achievement, other than in the broad sense for which they are designed, that being to make predictions of future academic success.

Since the inclusion of questions asking for literal and evaluative information activates variables such as passage organization and background information, tests attempting to measure these skills introduce complications into the assessment of overall reading ability. Reiterating earlier arguments, these skills include, besides reading ability, an ability to take tests, to process information hierarchically, to detect flaws, and possibly some others. A good reader, as we must understand the term from an objective assessment of standardized tests, is one who shows evidence of more than simply deriving information from writing. He or she is to some extent an expert problem solver, is well-informed, and sees things in the same way as the people who design tests. Whether or not this is how we should want to define a good reader is an educational issue on which the choice of modifying our testing strategies or keeping them as they are should be based.

The most important point to consider in deciding the issue is whether the tasks we now expect students to perform and the labels we hand out to students on the basis of test results have a stultifying effect on students. Do declared low achieving readers who, in spite of their labels, know they can extract information from writing, come to believe finally that they cannot read and so fulfill the predictions made about them? If so, then this is our strongest argument for more carefully considering the structural quality of the written materials we present to them and the real demands we make of them.

Bibliography

Baker, L. and N. Stein. 1981. The development of prose comprehension skills. In C.M. Santa and B.L. Hayes (eds.), Children's prose comprehension. Newark, DE: International Reading Association.

California Testing Bureau. 1978. California Achievement Tests: Class management guide, Forms C and D, Levels 10-19. Monterey, CA: CTB/McGraw-Hill.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (ed.), Subject and topic. New York: Academic Press.

Chafe, W. 1979. The flow of thought and the flow of language. In T. Givon (ed.), Syntax and semantics, Volume 12, Discourse and syntax. New York: Academic Press.

Clark, H.H. and S.E. Haviland. 1977. Comprehension and the given-new contract. In R.O. Freedle (ed.), Discourse production and comprehension. Norwood, NJ: Ablex.

Fillmore, C. 1974. Pragmatics and the description of discourse. In Berkeley studies in syntax and semantics, I. Berkeley: University of California, Berkeley, Department of Linguistics.

Fillmore, C. 1982. Ideal readers and real readers. In D. Tannen (ed.), Analyzing discourse: Text and talk. (Georgetown University Round Table on Languages and Linguistics 1981). Washington, DC: Georgetown University Press.

Frederickson, C. 1975. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology 7:371-458.

Grice, H.P. 1975. Logic and conversation. In P. Cole and J.L. Morgan (eds.), Syntax and semantics, Volume 3, Speech acts. New York: Academic Press.

Grimes, J. 1975. The thread of discourse. The Hague: Mouton.

Halliday, M.A.K. and R. Hasan. 1976. Cohesion in English. London: Longmans.

Hieronymus, A.N., E.F. Lindquist, and H.D. Hoover. 1979. Iowa Test of Basic Skills: Teacher's guide for administration, interpretation, and use, Levels 9-14. Iowa City: Houghton Mifflin.

Huggins, A.W.F. 1977. Syntactic aspects of reading comprehension. (Technical Report No. 33). Urbana, IL: University of Illinois, Center for the Study of Reading.

Keenan, E. and B. Schieffelin. 1976. Topic as a discourse notion: a study of topic in the conversations of children and adults. In C.N. Li (ed.) Subject and topic. New York: Academic Press.

Kintsch, W. 1977. On comprehending stories. In M.A. Just and P.A. Carpenter (eds.), Cognitive processes in comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Labov, W. and J. Waletzky. 1967. Narrative analysis: oral versions of personal experience. In J. Helm (ed.), Essays on the verbal and visual arts. Seattle: University of Washington Press.

Longacre, R. 1979. The paragraph as a grammatical unit. In T.Givon (ed.), Syntax and semantics, Volume 12, Discourse and syntax. New York: Academic Press.

Mandler, J.M. and N.S. Johnson. 1977. Remembrance of things parsed: story structure and recall. Cognitive Psychology 9:111-151.

Meyer, B.J.F. 1977. What is remembered from prose: a function of passage structure. In R.O. Freedle (ed.), Discourse production and comprehension. Norwood, NJ: Ablex.

Meyer, B.J.F. 1975. The organization of prose and its effects on memory. Amsterdam: North-Holland.

Ochs, E. 1979. Planned and unplanned discourse. In T. Givon (ed.), Syntax and semantics, Volume 12, Discourse and syntax. New York: Academic Press.

Olson, D.R. 1977. Open discussion on the contributions by Meyer and Carter. In R.C. Anderson, R.J. Spiro, and W.E. Montague (eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum.

Prescott, G.A., I.H. Balow, T.P. Hogan, and R.C. Farr. 1978. Metropolitan Achievement Tests: Teacher's manual for administering and interpreting. New York: The Psychological Corporation.

Rumelhart, D.E. 1975. Notes on a schema for stories. In D.G. Bobrow and A.M. Collins (eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press.

Rumelhart, D.E. 1980. Schemata: the building blocks of cognition. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D.E. and A. Ortony. 1977. The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, and W.E. Montague (eds.) Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum.

Sacks, H. and E. Schegloff. 1974. Two preferences in the organization of reference to persons in conversation and their interaction. In N.H. Avison and R.J. Wilson (eds.), Ethnomedology: Labelling theory and deviant behavior. London: Routledge and Kegan Paul.

Schank, R.C. 1975. The structure of episodes in memory. In D.G. Bobrow and A. Collins (eds.), Representation and understanding. New York: Academic Press.

Spiro, R.J. 1977. Remembering information from text: the "state of schema" approach. In R.C. Anderson, R.J. Spiro, and W.E. Montague (eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Lawrence Erlbaum.

Spiro, R.J. 1980. Constructuve processes in prose comprehension. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Stein, N.L. and C.G. Glenn 1978. An analysis of story comprehension in elementary school children. In R.O. Freedle (ed.), Discourse processing: Multidisciplinary perspectives. Norwood, NJ: Ablex.

Trabasso, T. 1981. Can we integrate research and instruction on reading comprehension? In C.M. Santa and B.L. Hayes (eds.), Children's prose comprehension. Newark, DE: International Reading Association.

van Dijk, T. 1977. Semantic macro-structures and knowledge frames in discourse comprehension. In M.A. Just and P.A. Carpenter (eds.), Cognitive processes in comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Vygotsky, L.S. 1978. Mind in society: the development of higher psychological processes. Cambridge, MA: Harvard University Press.

Winograd, T. 1975. Frames and the declarative-procedural controversy. In D. Bobrow and A. Collins (eds.), Representation and understanding. New York: Academic Press.

Winograd, T. 1977. A framework for understanding discourse. In M.A. Just andP.A. Carpenter (eds.), Cognitive processing in comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Young, R.E., A.L. Becker, and K.L. Pike 1970. Rhetoric: Discovery and change. New York: Harcourt, Brace, and World.